A Systematic Investigation of the Applicability of Datamining for the Improvement of Pedestrian Safety

by **Leila Cristina Muchanga**

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

DOCTOR OF PHILOSOPHY

Supervisor: Harry Evdorides **Co-Supervisor**: Miles Tight

College of Engineering and Physical Sciences

University of Birmingham

25.11.2021

Abstract

Extensive research has shown advanced technology as a promising component towards improving the safety of pedestrians in an urban road environment. In GB, in 2017, evidence from a national database on road data (i.e., STATS 19) summed 1305 road fatalities, with pedestrian fatalities representing c.a. 25.7% of the total road fatalities. Datamining, the technique adopted in this study, is a process used to convert raw data into useful information. Fewer studies, however, have investigated a systematic approach adopting datamining techniques (e.g., ML) to assess its level of quality and consistency. Therefore, the applicability of datamining to reveal undiscovered pedestrian road-crash casualty patterns more efficiently was further explored. This work uses ML to examine casualty statistics during a one-year period in GB. The investigation includes data regarding pedestrian attributes, the road environment and its physical infrastructure. A key advantage of ML over conventional statistics is that with minimal human intervention, advance warnings are given independently when exposed to new data. The adopted technique here, explores discrete classification rules. From a range of classification algorithms and for the road data given, the FCM was applied for its proven degree of objectivity and response sensitivity. To this end, the results anticipate pedestrian road-crash casualties mainly for elderly male pedestrians during the day with fine weather, with permanent speed limits set at 30 mph, particularly when neither general nor pedestrian facilities are made available. New pedestrian safety measures may be recommended as in, appropriate publicity, introducing variable speed limits, and further research focusing on developing non-existing pedestrian safety models on a number of pedestrian physical facilities (e.g., footbridge, pelican).

Keywords: *Discrete Classification, Machine-Learning, Data-Learning Algorithms, Unsupervised-Learning, FCM, Knowledge Discovery, Pedestrian Safety*

Acknowledgements

"A new journey to explore the unknown."

The author is greatly indebted to Dr. Harry Evdorides from the University of Birmingham for his considerable advice, support and encouragement to carry out this research project.

She is grateful to the University of Birmingham personnel, and especially thanks for their invaluable assistance in making the author familiar with the virtual facilities of the University.

Mrs. Jane Sjoberg and Mrs. Monika Davis are gratefully acknowledged for their useful comments at the writing stage of this thesis.

She thanks to Professor Emeritus Miles Tight from the Department of Civil Engineering for his co-supervision, advice and support at the initial and development stages of the project.

Contents

List of Figures

List of Tables

Acronyms

AIDS Acquired Immunodeficiency Syndrome. 2

APMs Accident Prediction Models. 85

BAC Blood Alcohol Content. 24

CART Classification and Regression Tree. 126, 127

CDT Classification Decision Tree. 127

CMF Crash Modification Factors. 84

DfT Department for Transport. 48

DST Danmarks Statistik. 3

DTANN Decision Tree Artificial Neural Network. 20

ENAY Number of Accidents Experienced per year. 91

EU STATS European Union Statistics. 4

FCM Fuzzy-Clustering-Means. ii, 128, 129, 134

GB Great Britain. ii, 6

GES General Estimates System. 14

GSRRS Global Status Report on Road Safety. 2, 3

iRAP International Road Assessment Programme. 8

KDD Knowledge Discovery Database. 72

KNIME The Konstanz Information Miner. vi, 64

MAPE Mean Absolute Percentage Error. 88

ML Machine-Learning. ii, 1

MLOps Machine-Learning Operations. 221

OECD Organisation for Economic Co-operation and Development. 3

PNAY Number of Predicted Accidents per year. 91

PRACT Predicting Road Accidents - a Transferable methodology across Europe. 14

RDT Regression Decision Tree. 126

ROC Receiver Operating Characteristic. 105, 110

SHAP Shapley Additive Explanations. 20

VRU Vulnerable Road Users. 3

WHO World Health Organization. 2

WRA World Road Association. 84

Glossary of Terms

C

clustering or cluster analysis is an unsupervised-learning task. 128

D

data-learning algorithms a technique used to extract appropriate knowledge for application in a new situation. 21

datamining a process used to turn raw data into useful information. ii, 1, 20

descriptive analysis a technique adopting ML to verify or train the data. 137

descriptor or train data, a description of a percentage of the data after a model has been applied. 59

discrete classification data sorted into distinct, countable, and in separate categories. These categories are converted into numeric variables with a countable number of values, with a limited amount of numbers. ii

F

FCM fuzzy clustering (also referred to as soft clustering or soft k-means) is a form of clustering in which each data point can belong to more than one cluster. ii

FCM demonstrator a partially-scaled approach to test the applicability of examining road-crash data from two different regions: GB, London. 137, 140

FCM prototype a fully-scaled approach to conduct with a spatial road-crash data investigation in GB. 134

G

gini calculation of the probability of a specific features classified incorrectly when selected randomly. 219

K

KDD or knowledge database discovery, similar to datamining, a process of discovering useful knowledge from collected data. 72

M

Machine-Learning a method of data analysis based on the idea that systems can learn new information data, identify patterns and make decisions with minimal human intervention. ii

O

overfitting when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.

219

P

partition in datamining is the division of the whole data available divided in percentages into two non-overlapping sub-modes: train, and test. 69

pedestrian facilities examined types of pedestrian physical facilities (e.g., zebra, pelican) wherein pedestrian road-crash casualty occurred. 14

pedestrian general attributes examined human factors influencing on pedestrian road-crash casualty causation . 5, 13, 201

pedestrian road-crash casualty a road-crash resulting in at least one pedestrian fatality. 5

pedestrian road-crash casualty causation reasons leading to a road-crash casualty fatalities. 72

predictive analysis a technique adopting ML to verify and test or validate the data. 123

predictor predictor or test data, a percentage of the data with an indicative projection of future events after a model has been applied. 59

prunning a data compression technique to remove redundant or the least important parts of a model or search space. 219

R

road elements road environment, road infrastructure influencing on pedestrian road-crash casualty causation. 7

road environment examined road environmental factors (e.g., speed limits, lighting condition) influencing on pedestrian road-crash casualties. 14

road infrastructure examined road physical infrastructural factors (e.g., road type, pedestrian physical facilities) influencing on pedestrian road-crash casualties. 14

road type examined types of road (e.g., intersection, roundabout) wherein a pedestrian road-crash casualty occurred. 14

road-crash when either a motorised or non-motorised vehicle collides with a pedestrian. 3

S

statistics percentage of actual pedestrian road-crash fatalities. 2

STATS19 is a code designating the protocol which outlines information to be collected whenever an injury crash is reported to the Police in the UK. 34

supervised-learning the data-learning algorithms have the task to learn a function that maps an input to an output based on input-output pairs. 11

U

unsupervised-learning it allows for data-learning algorithms to automatically discover natural grouping in data. Unlike supervised-learning enables finding natural groups in feature space. 11

urban road environment this nature of environments are directly associated the density of traffic, and the number of road users (e.g., pedestrians, drivers, bicyclists sharing the road). They can refer for example to towns, cities, and suburbs. ii

1. Introduction

1.1. Overall Approach

Existing research recognises the urgent need to investigate road safety challenges causing pedestrian casualties more in-depth. The past decade has seen increasingly rapid advances in the field of datamining to improve pedestrian safety. Datamining, the adopted approach in this study, is a process used to convert raw data into useful information (Wu et al., 2008; Prato et al., 2008; Addakiri et al., 2010; Giummarra et al., 2021. This research introduces a novel approach using datamining techniques (i.e., Machine-Learning (ML)) to explore undiscovered patterns and test the applicability of ML on a number of contributory factors that affect road-crash casualties (Butsenko, 2022; De Ona et al., 2013).

1.2. General Pedestrian Road Safety Assessment

There is a growing body of literature recognising the importance of pedestrian road safety. According to the World Health Organization (WHO), a major hindrance faced by road safety is that worldwide, every 23 seconds, a road user (e.g., pedestrian, cyclist, driver) dies. Looking just at pedestrians, the same happens every 101 seconds (WHO, 2018). WHO also states road safety as one of the greatest development challenges in the world, since the related human tragedy seriously impacts socioeconomical advancements (WHO, 2015; WHO, 2018). Worldwide and per annum, over 1.24 million people die on the roads. Moreover, it is estimated that between 2015 and 2030, 265 million people will be killed or seriously injured, surpassing Malaria, Acquired Immunodeficiency Syndrome (AIDS), tuberculosis, and diarrhoea diseases (WHO, 2018). Furthermore, unless urgent actions are taken, the number of people killed in road accidents per year is predicted to rise from 1.24 million to 2 million (WHO, 2013). Datamining seems to play a key role in preventing road accidents due to its ability to reveal unforeseen links from raw data (e.g., data monitoring) (VfR, 2015; Lin et al., 2019; STATS20). A fundamental question faced by many researchers is the proportional discrepancy in pedestrian safety from region to region, given that the reasons underpinning road accidents might be if not the same, at least similar [see Section 1.3]. Figure 1.1 illustrates the overall distribution of road-crash fatalities by road user, contrasting the world with Europe according to the Global Status Report on Road Safety (GSRRS) statistics (GSRRS, 2018).

Figure 1.1.: Road-crash casualty by road user type: World vs. Europe.

Source: *GSRRS, 2018*

Figure 1.1 shows the GSRRS statistics (GSRRS, 2018). The results show respectively the following road-crash fatality rates involving Vulnerable Road Users (VRU) (i.e., Worldwide vs. Europe):

i. Cyclists: 8% (i.e., 3% vs. 5%).

ii. Motorised (2-3 wheeled vehicles): 39% (i.e., 28% vs. 11%).

iii. Pedestrians: 50% (i.e., 23% vs. 27%)

1.3. Pedestrian Safety Overview

Investigating pedestrian road-crash casualties is a growing concern for road safety locally, regionally, and worldwide (DfT, 2019; Organisation for Economic Co-operation and Development (OECD), 2020). However, evidence indicates that road fatality rates tend to differ by type of road user and also by region (DfT, 2019; Danmarks Statistik (DST), 2020; OECD, 2020). For instance, of the road fatality rates (per million inhabitants) in Member States of the European Union (EU) in 2018, the lowest were represented by the UK (28), Denmark (30) and Ireland (31). By contrast, the highest were represented by Romania (96), Bulgaria (88), Latvia (78), and Croatia (77) (OECD, 2020). In addition to that, in 2019, the European Union Statistics (EU STATS) shows that in the EU, out of the whole road-crash fatalities, 21% consisted of pedestrians and 12% of cyclists (EU STATS, 2019). Moreover, road data from three distinct EU countries (i.e., GB, Denmark, and Romania) shows significant differences. The pedestrians vs. cyclists road fatality rates per country were as follows:

i. GB: 470 in total (27%) pedestrians vs. (6%) cyclists (DfT, 2019).

ii. Denmark: 30 in total (15%) vs. 31 (16%) (DST, 2020).

iii. Romania: data from 2019 not found. In 2018, 37% vs. 9% (OECD, 2020).

In 2019, the key statistics database of EU countries reported that Romania had one of the highest fatalities in road accidents in all of Europe (Eurostat, 2019). Along with this, recent research work conducted by Sârbescu et al. (2014) and Hamann et al.(2015), indicated that in Romania the main causes of road-crashes were driver related (e.g., speed driving) or due to pedestrian jaywalking (Hamann, et al., 2015). Previous research concentrating on the costs of road-crashes has pointed to a lack of relevant information. Thus, the improvement of a system to collect and examine road data more exhaustively has been proposed (Drosu and Cofaru, 2016; Cadar et al., 2017). Concerning GB pedestrian road-crash casualties, the central country this study referred to, it shows a 25% decreasing tendency between 2005 and 2009. By contrast, further analysis shows a 12% increasing tendency between 2009 and 2015 (DfT, 2017). Still, describing reasons underpinning the fluctuation rates on pedestrian safety with a decade of data (i.e., 2005–2014) remains restricted (DfT, 2019). One of the reasons could be, among others, due to information corruption through omission, overwriting, or even misperception (Noland et al., 2017; Li et al., 2017). Problems of this kind may be ongoing. To address this issue, advancing cross-sectional studies to draw more conclusive results has been previously suggested (Urie et al., 2016; Useche et al., 2021). Besides that, some research with specific interventions to improve pedestrian accessibility was previously carried out (Jones and Anciaes, 2018; Lin et al., 2019). However, there remains a paucity of evidence on the pedestrian road-crash causal factors. This might likely be improved by, for example, correlating data from pedestrian general attributes (Uzan and Wagstaff 2017; Hezaveh and Cherry, 2018), the road environment (Sullivan and Flannagan, 2002 ; Naik et al., 2016; Malin et al., 2020) and the physical road infrastructure (Pardillo Mayora and Jurado Piña, 2009; Zegeer and Bushel, 2012; Gross et al., 2013; Vijayawargiya and Rokade 2017). Moreover, the adoption of modelling has proven its use to examine road data more efficiently (Stone et al., 2002; Nilsson 2004; Wedagama and Dissanayake, 2010; Zegeer et al., 2012; Aziz et al., 2013; Kaplan and Prato, 2012; Papadimitriou et al., 2019). Even so, the use of systematic methods to assist in reducing pedestrian road-crashes, remains unclear. To this end, the thesis adopts a systematic investigation to determine the extent to which a pedestrian safety modelling approach can more objectively assess pedestrian road-crash casualty trends. This will be shown in Chapters 7 and 8. Here, it is also of importance to mention that analysing data regarding drivers, cyclists, or motorcyclists is out of the scope of this study.

1.3.1. Pedestrian Road Fatalities in the UK

In the last decade, in Europe, pedestrian road-crash casualty rates have decreased by ca. 22% (Eurostat, 2019). On the other hand, the UK shows a decreasing tendency of ca. 3% (DfT, 2020). However, to date, the reasons for the existing gap between the two regions have not been fully clarified. One could argue that comparably, the subject complexity could increase based on the number of fatality rates by region (DfT, 2020; OECD, 2020). More particularly in the UK, looking at the Great Britain (GB) pedestrian road-crash casualties, which essentially regards the analysed data in this study, between 2005 and 2009, a marked decreasing tendency of 27% was noticed. In contrast, between 2009 and 2015, a slight increasing tendency of 4% was noticed (DfT, 2017). However, exceptionally from 2013 to 2014, a steep increasing tendency of 12% was noticed (DfT, 2015). Figure 1.2, illustrates a general overview of the mortality of pedestrians hit by vehicles in the entire UK.

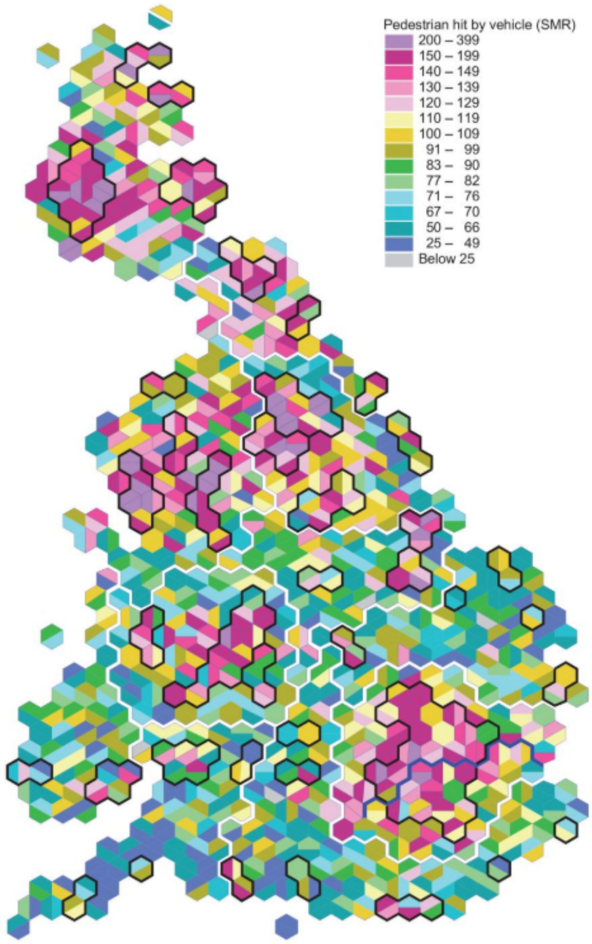


Figure 1.2.: UK pedestrian road-crash casualty overview

Source: *Shaw et al., 2008*

Figure 1.2 presents a general overview of UK pedestrian road-crash mortality distribution rates. However, depending on the pedestrian behaviour, the road elements, and the respective region, the nature of road-crashes is expected to differ (Stefanova, 2015). One critical drawback is, failing to understand pedestrian road-crash patterns (e.g., distribution, rates, and variation), as this may increase road-crash casualty causation. Another critical aspect is when the nature of these causes remains unclear. However, the number of systematic investigations on pedestrian safety using datamining is limited. Therefore, this study assesses the applicability of modelling to systematically examine road-crash data combining pedestrian general attributes, the road environment, and the road infrastructure.

1.3.2. Datamining Techniques

Essentially, datamining is a process available to databases used to convert raw data from multiple rows and columns of data into useful information (Butsenko, 2021). In the context of road safety, although not systematically, an extensive variety of datamining techniques have been previously applied (Sohn and Shin, 2001; Chang and Wang, 2006; Silva et al., 2020). Moreover, these techniques have successfully helped reveal patterns from road accident databases (De Oña López et al., 2013; Martin et al., 2014). ML is the datamining technique chosen in this study. This is because, when compared with traditional statistical techniques, it demonstrates more flexibility. An example is its ability to indicate patterns with a lower dependency on assumptions (Gu et al., 2017; Gutierrez-Osorio and Pedraza, 2020).

1.4. Problem Definition and Originality

A number of factors affecting pedestrian safety have been explored in several studies. Among others, these included pedestrian general attributes, the road environment, and the road infrastructure (Salmon et al., 2016). However, previous studies demonstrate limitations in showing the contribution of systems towards pedestrian safety (Pilkington and Kinr, 2005). This is because examining a wider range of pedestrian attributes and road attributes with a singular analysis may play a vital role in improving pedestrian safety (Urie et al., 2016; Jones and Anciaes, 2018; Lin et al., 2019 ; Useche et al., 2021). Regarding pedestrian general attributes in a specific area (e.g., kindergarten, schools), Zhu et al. (2021) present inconsistencies regarding the proportion of young pedestrians in relation to their road-crash risk propensity. As regards the road environment, Clarke et al. (2006) claim the time of day impacts both the severity and the number of road-crashes. Additionally, a recent data analysis shows road-crash rates with rainy weather increasing up to 71% (Khan et al., 2020). With respect to the road infrastructure, a previous study carried out in Sweden indicates the condition of the road to at least mitigate 59% of the fatal road-crash rates (Stigson et al., 2008). Additionally, the International Road Assessment Programme (iRAP) ranks 5 star roads as the safest in contrast with 1 star roads as the most risky (iRAP, 2021). Still, more than 50% world's roads are ranked as 1 or 2 stars (iRAP, 2015). As regards the improvement of systems to reduce pedestrian road-crashes, recent studies have used advanced data analysis systems (e.g., predictive modelling) (Deb et al., 2017). As a result, modelling rapidly became a key instrument in examining road data, for its lower dependency on historical data and also for minimising human intervention (Gu et al., 2017; Li et al., 2021) [Section 1.3]. However, in-depth data analysis that examines pedestrian general attributes in a road environment has not been addressed in much detail. One critical drawback of combining several attributes, when using either similar or different features on a single analysis, is associated with the model expiry (PRACT, 2016). This is because, it may contribute to the increase in important decisions being made intuitively, rather than evidence based (e.g., rich data stored in databases) (Han and Kamber, 2012). Three other possible limitations of a data-driven method are, high dependency on historical data, a lack of standardised data sources, and human intervention (Deb et al., 2017, Assi et al., 2020). These factors, seem to significantly weaken the relevance of the conclusions (Salmon et al., 2016; Nashad et al., 2016; Takahashi al., 2020). Therefore, to assist policy-making, engineering planning, and implementation, defining an adaptive data-driven procedure to constitute and implement a standard, a specification, or a regulation might become crucial (WHO, 2015).

1.5. Aim

The aim of this study is to improve pedestrian road safety, by assessing the applicability of an advanced data-driven method to systematically examine road-crash casualty causation involving pedestrians.

1.5.1. Objectives

i. To identify features that affect pedestrian road-crash casualties in relation to pedestrian general attributes, road environment, and road infrastructure.

ii. To use data from STATS 19 and explore ML algorithms to describe and project safety trends for pedestrians.

iii. To define a knowledge discovery approach for a safer pedestrian road environment.

iv. To test the applicability of the above methodology using data from GB and from different regions.

v. To review the literature related to road safety worldwide.

vi. To provide recommendations on road safety based on the data analysis and validate the findings with expert opinion.

1.5.2. Thesis Layout

The thesis details a data-driven study conducted during this Ph.D. For this reason, the thesis structure is composed of 11 themed chapters.

Chapter 1 introduces the fundamental elements to develop the overall investigation purpose.

Chapter 2 presents the background work, that includes exploring the applicability of relevant data tools and techniques to examine road data, and also concluding existing gaps.

Chapter 3 sets out the adopted methodology for this study.

Chapter 4 investigates previously applied predictive models, and gives focus to pedestrian attributes affecting pedestrian safety.

Chapter 5 assesses previously applied predictive models, and focuses on road elements impacting on pedestrian safety.

Chapter 6 explores the process of examining road data.

Chapters 7 details data-learning algorithms: supervised-learning supervised-learning, and unsupervised-learning.

Chapter 8 rationalises the effects of the applied unsupervised-learning algorithms on preassessed road data.

Chapter 9 tests the sensitivity response of the applied data learning algorithm, discusses the reliability of the results, and suggests possible measures to improve pedestrian safety.

Chapter 10 discusses the findings of the applied unsupervised-learning algorithms, and proposes improvements to examine road safety information.

Chapter 11 summarises the findings and presents some pedestrian safety recommendations, including practical applicability and future work.

2. Literature Review

2.1. Introduction

This chapter reviews the worldwide literature on existing pedestrian road-crash models to improve road safety. The significance of examining a number of contributory factors to pedestrian road-crash casualties is central to this study. This chapter is structured in six sections. Section 2.1 introduces the overall structure of the chapter. Section 2.2 distinguishes between two types of pedestrian safety models: traditional and predictive. Section 2.3 describes some of the influential factors on pedestrian safety as pedestrian attributes. Section 2.4 expands knowledge on physical infrastructure factors affecting pedestrian safety in an urban road environment. Section 2.5 explains the applicability of two types of pedestrian safety models previously applied to pedestrian road-crash data: supervised-learning, and unsupervised-learning. Section 2.6 summarises the findings and identifies pedestrian safety limitations, pointing to further research.

2.1.1. Overall Approach

Previous research has established that road data can help prevent road crashes (Noland et al., 2017). Moreover, Cahill (2010) and Zangenehpour et al. (2015) have previously suggested investigating further cross-sectional methods to improve pedestrian safety. However, much of the research seems unable to encompass exploring undiscovered patterns in more detail. In this study, ML was adopted to identify previously undiscovered links between pedestrian general attributes (Nemire et al., 2016; Earl et al., 2018; Hezaveh and Cherry, 2018; Feliciani et al., 2020) and road elements (Elvik. et al., 2009; Godavarthy and Russell, 2016; Noland et al., 2017; Oviedo-Trespalacios et al., 2017). Moreover, to discover new knowledge regarding their effects on pedestrian safety. This research sought to demonstrate how advanced data analysis, such as descriptive modelling, can be used to extract knowledge from a dataset. The data analysis gave focus to the application of ML using computational models without focusing on developing new models, let alone programming them.

2.2. Road-Crash Pedestrian Safety Models

Beshah and Hill (2010) hold the view that appropriate measures can prevent pedestrian road-crash casualties. Moreover, recent advances using datamining methods have proven their applicability to reduce road-crashes [see Chapter 7]. However, it is important to note that the relevance of existing data-driven methods seems to quite likely expire (Kaplan and Prato, 2012, PRACT, 2016, Meyer, 2021). Forecasting (i.e., classical) and predictive models (i.e., ML) are used to make predictions on future events (Torbic et al., 2010; Feliciani et al., 2017). Their differences are mainly dependent on the use of specific applications, methods, models, and on their level of accuracy in the predictions. Moreover, the accuracy of predictive modelling constitutes an important assessment measure, particularly when facing uncertainty (e.g., pedestrian behaviour) and wider-scaled road environments (e.g., city, country). This is because it can lead to provisioning erroneous information (Abdar et al., 2021, Maryam et al., 2023). This section presents two types of methods: classical and ML. The term classical data analysis refers to a number of models used to analyse factors affecting pedestrian safety through traditional statistical methods. Two examples are the mathematical model deployed by Nilsson (Nilsson, 2004) and the negative binomial regression for intersections (Torbic et al., 2010). This will be further explored in Chapters 4 and 5. The ML data analysis adopted in this study incorporates algorithms to jointly examine pedestrian safety causal factors and describe undiscovered safety trends. This will be expanded in Chapters 6 and 7.

2.2.1. Traditional Pedestrian Safety Models Assessment

A considerable number of models have been previously incorporated within the "Predicting Road Accidents - a Transferable methodology across Europe (PRACT)" database to analyse factors affecting pedestrian safety (PRACT, 2016). Generally, this type of study explores unforeseen pedestrian road-crash trends based on analysing the impact of elements such as pedestrian general attributes (i.e., gender, age group), the road environment (e.g., weather, speed limits, lighting) and the road infrastructure (e.g., road type, pedestrian facilities). However, fewer studies suggest a systematic association between these elements. Regarding pedestrian general attributes, in the USA, between 2005 and 2009, data on gender and age group from the National Automotive Sampling System "General Estimates System (GES)" was used (Kaplan and Prato, 2012). One drawback is the limited data of 1% on road accidents for composing prediction samples. This becomes a key issue for scalability purposes (Trueck and Rachev, 2009). Another study carried out in 2015 in Italy, regarded a predictive simulation comparing the behaviour between the young and the elderly in a road environment. However, the simulation did not assess their movement by gender (Zeng et al., 2014; Gorrini et al., 2018). Regarding the road environment, an example is the integer autoregressive model, which was adopted to examine the influence of the weather in Athens, Greece (Yannis and Karlaftis, 2010). This analysis has neither dealt with incorporating various models nor with the lack of independence from historical data. A further explanation of models to examine speed limits and lighting will be expanded in Chapter 5. On the road infrastructure, at an intersection, when compared with more recent models (Prato et al., 2012), the negative binomial regression model also requires large data sets (e.g., years, number of locations) (Lyon and Persaud, 2002; Torbic et al., 2010). Additionally, further research on pedestrian safety at signalised intersections was suggested (Zegeer et al., 2005). On pedestrian physical facilities, a mathematical formula was used to examine the effects of their absence and presence (iRAP, 2018; Gitelman et al., 2012). Similar to the road environment, models regarding other types of road infrastructure will be expanded in Chapter 5. Despite the progress made to date to improve pedestrian safety, the literature review identified existing research gaps as follows:

i. Human attributes: According to Oström and Eriksson (2001), in a road environment, pedestrians initiating collisions have been mostly descriptive (e.g., time and place) and epidemiological (i.e., factors medically diagnosed that determine road-crashes as a disorder) rather than exploratory (i.e., knowledge discovery driven). For instance, a pedestrian classified by gender and age group that is permanently impaired (e.g., cognitive, physical) indicates a significant share of unpredictability (Guth et al. 2012; Castillo-Manzano et al., 2017;

Uzan and Wagstaff, 2017). ML can be adopted to simulate unforeseen movements (e.g., jaywalking) of pedestrians with a level of impairment in relation to existing pedestrian physical facilities (Feliciani et al., 2017). However, an enhanced approach that accounts for this issue has not been explored to date. This research explores ML to overcome the limitations of existing methods by gender and age group. Subsequently, the newly discovered knowledge can play a critical role in assisting with recommending relevant physical infrastructure that accounts for pedestrian attributes such as gender and age group (e.g., children, elderly).

ii. Road environment factors: Poor weather is perceived as a barrier to physical activity, and it may become an issue, especially for impaired pedestrians (Theofilatos et al., 2014). One critical issue is the low accuracy of models in terms of weather forecasting (Naik et al., 2016). ML can be adapted to assist in the implementation of physical infrastructure that accounts for the weather influence (e.g., rainfall, snow) (Ryser and Halseth, 2019). Regarding speed limits, in São Paulo, Brazil, a study from 2016 shows variable speed limits leading to diverging road-crashes to alternative routes (Ang et al., 2020). One critical limitation of this examination is the degree of imprecision of speed limits impacting road-crashes (Granà et al., 2010; Bichicchi et al., 2017). ML can assist in projecting speed limit measurements to rate the level of safety by route (Peng et al., 2022). With respect to lighting, it seems to significantly increase pedestrian safety, especially at night (Fotios and Gibbons, 2018). One issue is the significant amount of unknown illumination at night (ca. 25.5%). ML can be adopted to implement adaptive street lighting (Tripathy et al., 2017). Furthermore, to date, ML applications have not clearly dealt with assessing the models' accuracy after combining multiple factors affecting pedestrian safety (e.g., weather, speed limits, and lighting) on a single examination. This research assesses the accuracy rates of an ML algorithm associated with multiple factors (e.g., weather, speed limits, and lighting) with the view to providing relevant pedestrian safety recommendations.

iii. Road infrastructure: Differences based on the type of intersections (e.g., unsignalised, signalised, and roundabouts) and depending on their layout (e.g., geometric design, traffic signage) directly impact on road-crashes (Obeng and Rokonuzzaman, 2013). ML can support defining more relevant road layouts through simulation environments (Feliciani et al., 2017; Onelcin and Alver, 2017). However, an enhanced approach to examine road-crashes in relation to the infrastructure layout has not been made to date. This research explored further the applicability of ML to address current setting constraints.

v. Data quality: Lack of data consistency or data completion (e.g., human failure) may disqualify the validity of projecting interactions between different elements in a road environment (Nashad et al., 2016; Schmidt et al., 2021). Moreover, an enhanced approach using ML to minimize the historical data dependency has not been made to date. In contrast with previously adopted models (Torbic et al., 2010), a key advantage of incorporating ML, lies in its capacity to define patterns with a reduced reliance on historical data (Prato et al., 2012). This research specifically applied ML to data collected from a single year, aiming to demonstrate a reduction in historical data dependencies.

Therefore, the ML approach seems to enhance the current needs of pedestrian safety research by means of:

i. Knowledge discovery methods that have been applied to examine unforeseen pedestrian movement.

ii. Development of pedestrian safety models that can adequately describe the influence of more than one factor associated with the road environment.

iii. Development of models that can satisfactorily describe the influence of data quality in spatially diverse datasets.

2.2.2. Conventional Approach

Figure 2.1 illustrates a classical data analysis concept to analyse pedestrian road-crash casualty data [see Section 2.2].

Figure 2.1.: Classical data analysis

The conventional analysis displayed in Figure 2.1 includes the following four stages: pre-definition of the elements, data analysis, reporting, and forecasting (DfT, 2017). About the elements pre-definition, a number of elements (e.g., human, road environment, and road infrastructure) are suggested (Urie et al., 2016; Jones and Anciaes, 2018; Lin et al., 2019; Useche et al., 2021). Traditionally, prior to initiating a classical analysis (e.g., forecasting through human calculation), the data goes through a cleansing process (e.g., data filter) (DfT, 2017). As regards reporting, the results are expected to be incorporated into a computational datamining tool. These findings are expected to clarify pedestrian road-crash casualty causation trends in relation to the selected elements, such as pedestrian attributes (e.g., impairment level due to alcohol or drugs consumption, and road elements (e.g., speed limits, single carriageway, footbridge) (DfT, 2017). Forecasting refers to traditional statistical models (e.g., logistic regression, binary regression) to project future events. Two identified major issues that might likely affect the quality of the analysis, are the high dependency on past events and human intervention. Regarding the high dependency from data, typically this type of analysis utilises large sets of historical data (DfT, 2017; OECD, 2020; Sabrina, 2020). Since this nature of data also depends on human intervention (e.g., misspelled records, new terminologies introduction), data inaccuracies or incompleteness are expected (Noland et al., 2017) [see Chapter 2]. A consequence of discarding defects when forecasting data could be the disqualification of the findings, for the region under investigation and when attempting to transfer a common approach to different regions (Imprialou and Quddus, 2019; Noland et al., 2017; Feliciani et al., 2020; Giummarra et al. 2021). Therefore, to investigate pedestrian road-crash in GB with a lower dependency of data, a pre-eminent data-driven method is proposed.

2.2.3. Machine-Learning for Pedestrian Safety Prevention

ML has been developed to learn and define undiscovered patterns of any nature of data (Trueck and Rachev, 2009; Witten et al., 2011; Addakiri et al., 2019). Essentially, ML distinguishes between two different types of learning techniques: supervised and unsupervised. The supervised-learning maps data added in a computational datamining tool, i.e., data feed, to a result based on an example of a data feed, data result pairs (Sanz-Casado et al., 2019; Assi et al. 2020). The unsupervised-learning seeks out undetected patterns, with a more reduced amount of data feed, and also with a minimised human intervention (Feliciani et al., 2020; Giummarra et al. 2021). Compared with traditional modelling techniques, ML algorithms applied to pedestrian safety have become popular because of their flexibility to be used by humans (KNIME, 2021). Moreover, ML models have been used to identify trends by integrating a wider range of contributing factors (e.g., pedestrian attributes, pedestrian facilities). Subsequently, the provision of firmer recommendations towards pedestrian safety becoming more objective and also more accurate (Max and Johnson, 2013; Akash et al., 2014; PRACT, 2016). Turning to ML applied to pedestrian safety, the results from a study carried out during an 11-year period (2006–2016) in Colorado, USA, presents the extreme Gradient Boosting (XGBoost) model to reach up to 89% in accuracy. The XGBoost was used to classify three different levels of severity in terms of pedestrian road-crash data for the elderly. However, to more accurately interpret the data, the XGBoost required to include the Shapley Additive Explanations (SHAP) technique (Guo et al., 2019; Li et al. 2020). Another study carried out in the USA, during a 6-year period (1995–2000) to examine the speed limits influence on road-crashes applied a hybrid Decision Tree Artificial Neural Network (DTANN) model. For fatal injuries, the results show DTANN training and testing performance of 91.53% and 90% respectively. However, disappointingly, most of the data records on speed limits (i.e., 67.68%), were unknown. Therefore, it is believed that a plausible measure of improvement could include accessing a greater amount of data (Chong et al., 2004).

2.2.4. Valuation of the Models Applicability

Recent studies using datamining indicated significant advancements towards pedestrian safety. This will be presented later in Chapter 7. However, initial observations indicate links between pedestrians and the infrastructure to remain unperceived [see

Sections 2.2.1, 2.2.3]. In addition to that, the applicability of the currently used data-learning algorithms may likely become outdated (PRACT, 2016). Moreover, fewer models prove to adopt common variables to examine distinct types of road-crashes [see Chapter 1; Section 2.2.3]. This indicates that the findings cannot be extrapolated to all road-crashes. Subsequently, in an attempt to provide a recommendation or implement a policy, challenges might quite likely arise. Furthermore, it is of equal importance to highlight that, to enhance the applicability of ML to explore undiscovered pedestrian safety patterns, data quality plays a dominant role. Therefore, contributory factors such as pedestrian attributes, the infrastructure, and the quality of the data are fundamental to the current discussions on pedestrian safety. This will be expanded in Chapters 4], 5], and 6. Even though this study does not directly engage with additional factors other than pedestrian general attributes (e.g., drivers, riders) and all road elements (e.g., special conditions on site, carriageway hazards), a more comprehensive approach could be explored. The exploitation of these factors is out of the scope of this study.

2.3. The Influence of Human Factors on Pedestrian Safety

A pedestrian while commuting may not strictly follow rules, especially those diagnosed with cognitive impairment (Nhac-Vu et al., 2014; Cordellieri et al., 2019; Jagnoor and Peden, 2020; Giummarra et al., 2021). Additionally, factors such as perception and distraction were considered highly influential (Bungum et al., 2005; De Oxley et al., 2005; Schuurman et al., 2009; Jones and Anciaes, 2017). For example, in some cases, pedestrians may decide on a route based on what they consider sensible and necessary to arrive at their desired destination. Evidence from the USA between 1995 and 1998 shows 63% of pedestrian road-crashes occur when crossing the road (da Silva et al., 2003). The authors also state that, illegally crossing the street may be the main reason for a pedestrian road-crash to happen. In addition to that, previous research studies indicate pedestrians distracted by telephone conversations or other activities (e.g., eating, listening to music) are at greater risk when crossing the street (NHTSA, 2009; WHO, 2018). Gariazzo et al. (2018) reported road-crashes resulting in fatalities when a pedestrian was talking on a mobile phone or using an I-pod (Mwakalonge et al., 2015). Pedestrian safety intervention programs have been previously suggested (De Oxley et al., 2005; Schuurman et al., 2009; Jones and Anciaes, 2017).

2.3.1. Pedestrian General Attributes

The pedestrians scored the highest rank on road-crash exposure [see Section 2.3]. A number of pedestrian general attributes affecting pedestrian road-crash casualties are illustrated in Figure 2.2.

Figure 2.2 displays the pedestrian general attributes categories: gender and age group. As regards gender, it is classified as females and males With respect to the age group, it is classified as follows: children (0–15 years old), young adolescents (15–24 years old), adults (25–59 years old), and elderly (60+ years old). Apparently, human error highly affects the propensity of pedestrian road-crashes (Ma et al., 2010; Zheng et al., 2019). For example, on risk responsiveness, results from a previous study conducted by Zheng et al. (2017), show personality traits associated with risk perception and acceptance leading to specific behaviours. Figure 2.2 shows the nature of pedestrian general attributes further observed through case studies: risk perception, distraction, and impairment (Allen and Singh, 2011; Dultz and Frangos, 2013; Ayers et al., 2016; Zhang et al., 2019).

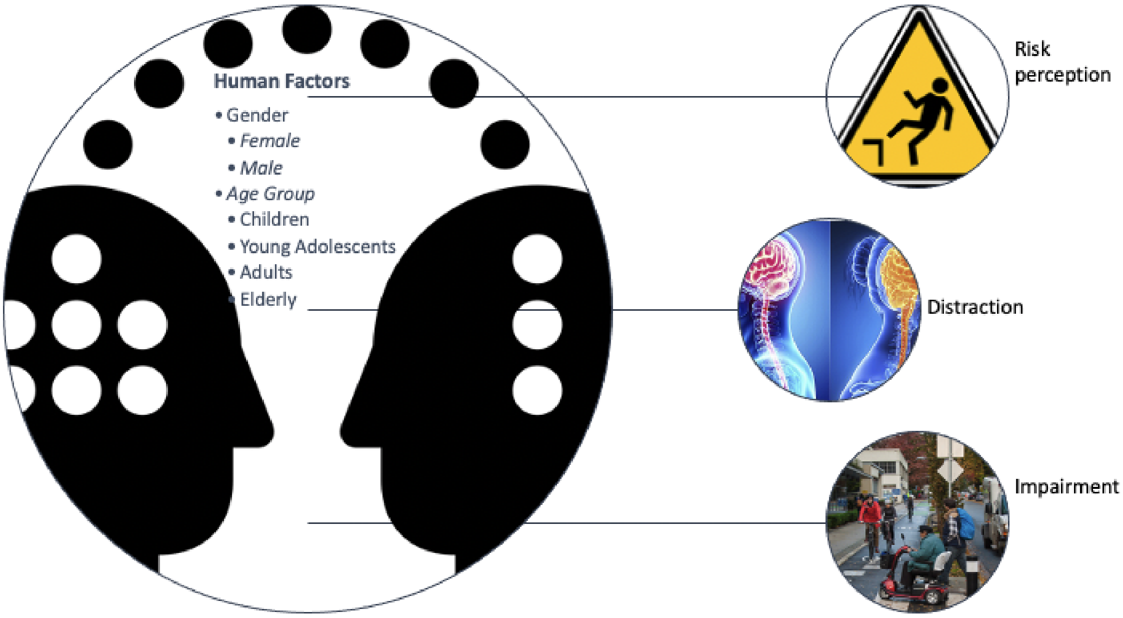


Figure 2.2.: Explored pedestrian general attributes

Pedestrian Risk Perception

A theoretical distinction between cognitive judgment (e.g., risk perception) and the risk-related worry (e.g., concern about the risk consequences) has been previously proposed (Sjöberg, 2006; Cordellieri et al., 2019). Regarding the lack of cognitive judgement, specific interventions to improve accessibility such as adding or improving ramps have been previously recommended (Jones and Anciaes, 2018). However, the authors draw attention to a potential pedestrian safety challenge: the dynamics of distinguishing between risk perception and worry (Cordellieri et al., 2019). Therefore, preventive measures as provision of safe road environments and development of behavioural and training packages have also been proposed (De Oxley et al., 2005; Schuurman et al., 2009).

Pedestrian Distraction

Pedestrians are likely at a greater risk when crossing the street mainly due to distraction by telephone conversations or other activities (e.g., eating, listening to music) (Bungum et al., 2005; Syazwan et al., 2016). In addition to that, mobile phone conversations seem to induce cognitive distraction (Tornros and Boiling, 2005; Hou et al., 2020). Interestingly, the same authors propose to introduce warnings about driving through games, to anticipate road-crashes involving both pedestrians and drivers (Ayers et al., 2016). It is important to note that these findings should be interpreted cautiously and not simply extended to the general population (Hatfield and Murphy, 2007). Nevertheless, authorised laws to restrict hand-held devices for telephone conversations or texting have proved to contribute to reduce pedestrian road-crashes (Huang et al., 2010).

Pedestrian Impairment

Various research studies demonstrate pedestrian road-crashes involving road users under the influence of chemical substances hold a significant share in fatalities (Demetriades et al., 2004; Castillo-Manzano et al., 2017). According to Elliott et al. (2009) and Treno et al. (2014), there are two main types of chemical substances seriously affecting pedestrian safety. Concerning the consumption of drugs, Mura et al. (2006) agree by claiming these (e.g., cannabis plant) to seriously influence pedestrian road accidents. Regarding the consumption of alcohol, evidence from a case study indicates that pedestrians highly intoxicated by alcohol (i.e., Blood Alcohol Content (BAC) 0.07–0.10%) seem to have impaired judgement and to be under performing physically when crossing the road (Clayton and Colgan, 2001). In addition to that, the pedestrians tend to expose themselves to more severe injuries (Dultz and Frangos, 2013). Another case study carried out in Tennessee, between

2011 and 2016, identified that 7% of the road-crashes involved pedestrians walking under the influence of alcohol (Hezaveh and Cherry, 2018). Turning now to a more permanent nature of impairment, two main types of disability will be further described: physical and cognitive. With respect to physical disability, the results from a study conducted by Guth et al. (2012) in a residential area of Tampa, Florida, USA, reveal the blind (or nearly blind) to more frequently be at risk when crossing the street, especially in a high traffic volume setting. In this location, the drawback to redesign crosswalks adapted to the general population of blind users was the restriction on the number of participants (19 in total: ten blind or nearly; nine normal vision) (Uzan and Wagstaff, 2017). About cognitive disability, Earl et al. (2018) explored the behaviour of pedestrians with and without a cognitive impairment when getting close to pedestrian facilities (e.g., zebra crossing). The results categorise two groups: confident users and users who “know what -they- are doing but drivers might not”. One key limitation was to diagnose the degree of impairment of participants (e.g., ethical, practical). In overall, in terms of the impairment both the cognitive (e.g., drugs, alcohol, dementia) (Elliott et al., 2009; Treno et al., 2014; Earl et al., 2018; Hezaveh and Cherry, 2018) and the physical (Guth et al. 2012; Uzan and Wagstaff, 2018) share a common weakness: these cannot objectively measure the danger impact of the different impairment levels.

Table 2.3.1, synthesises the findings from distinct observational studies regarding pedestrian road crossing behaviours.

Table 2.1.: Pedestrian walking speeds: observational behaviour

Observed behaviours	**Results**	**References**
Surroundings awareness	*Yes*(%)	
Risk Perception		(Bungum et al., 2005)
Waited for a walk signal	58.2	
Entered walk after yellow light	24.6	
Distraction		(Bungum et al., 2005)
Eating, Drinking, Smoking	15.1	
Headphones, Mobile phone	5.7	
Impairment influence	*Findings*	
Walking speed of shorter term disabled pedestrians:		
Alcohol, drugs	Data not found	
Walking speed of longer term disabled pedestrians:		
Physical	0.87m/s	(Awaludin et al., 2021)
Cognitive	Data not found	
Studies on the effects on walking due to cognitive impairment as:		(Oxley et al., 2005)
Dementia	1 out of 1	
Medical conditions	10 out of 10	

As per Table 2.3.1, the observational study conducted between 1998 and 2000, by Bungum et al. (2005) in Las Vegas, Nevada, USA, measures the pedestrians perception and their distraction levels when crossing an intersection near to a university. From this study it was found that 20% of the pedestrians were distracted as they crossed the street. Subsequently, inexpensive education training targeted at pedestrians near college campuses was recommended. However, disappointingly, data on walking speeds for pedestrians with different cognitive disabilities was not found. Moreover, some other studies state limitations to measuring the impact on road crashes when taking into account pedestrians with different cognitive levels (Treno et al., 2014; Hezaveh and Cherry, 2018). Nevertheless, even though the summarised results indicate pedestrians to most likely be cautious when crossing the road, a limited body of literature on pedestrian cognitive behaviour was found (Oxley et al., 2005; Jomnonkwao et al., 2021). Subsequently, extrapolating the applicability of the aforementioned factors to different realities (e.g., human cognition, pedestrian location) or temporary circumstances (e.g., rush hour, pedestrians under alcohol influence), could become a crucial challenge when attempting to examine pedestrian safety trends.

2.3.2. Gender

Road-crash casualties are more prevalent with males than females. Moreover, it has been previously found that female pedestrians tend to be more cautious than male (Zhu et al., 2012; Rankavat and Tiwari, 2016; Onieva-García et al., 2016). Females are apparently less likely exposed to danger on the roads (Onieva-García et al. 2016; Useche et al., 2019). However, unforeseen events might supersede the trends. A previously conducted study showed an unexpected traffic intensity increase resulting in a steep rise in female pedestrians crossing red light traffic signs (Ren et al., 2011). Regarding the consumption of alcohol in relation to pedestrian road-crashes, these tend to rise when the in-taken of ethyl alcohol increases. In addition to that, on the levels of alcohol intake, males were prevalent over females (Hezaveh and Cherry, 2018; Lasota et al., 2019). Moreover, it has been found that males are nearly twice as much at risk as females (Useche et al., 2019). One reason could be that male pedestrians tend to be less sensitive to traffic safety (e.g., faster walking) (Holland and Hill, 2007). Another reason could be their behaviour under the influence alcohol, which increases pedestrian road-crashes significantly (Hezaveh and Cherry, 2018). In overall, the results (i.e., males/females) are respectively represented by the following rates:

i.Canada: 55%/45% (PDR, 2010).

ii.Europe: 64%/36% (TSFS, 2018).

iii.UK: 56%/24%/ (DfT, 2020).

A study carried out in Karachi, Pakistan, shows males to more likely than females be prone to jump off, get on, and run to catch a moving bus (WHO, 2003). Another study in the USA, between 2008 and 2009, also shows males being exposed to higher road-crash casualty rates (Zhu et al., 2012). In general, various studies showed similar results, with the difference that exposure rates for females fluctuated (Onieva-García et al., 2016). Table 2.2 shows pedestrian walking speeds by gender.

Table 2.2.: Walking speeds by gender

Walking speed by Gender	*Mean(m/s)/SD*	*Skewness Kurtosis*
Single Carriageway		
Female	1.42/0.30	0.7148 -0.1422
Male	1.52/0.28	1.0086 1.4912
Double carriageway		
Female	1.45/0.23	0.9512 1.8067
Male	1.47/0.24	0.5000 -0.6700

Source: (Montufar et al., 2012; Chandra and Bharti, 2013)

The results in Table 2.2 show significant differences in walking speeds depending on the type of facilities (i.e., single carriageway, double carriageway). On average, the walking speed is higher for the male when crossing single carriageways. Interestingly, on double carriageways the walking speed increases for the female and decreases for the male (Wilson and Grayson 1980; Montufar et al., 2012). In this case, the female demonstrated a higher degree of risk perception (Rankavat and Tiwari, 2016; Onieva-García et al., 2016) and also to expose themselves less to risk (NHTSA, 2010; Zhu et al., 2012). Moreover, the results show a positive skewness for both single carriageways (i.e., right skewness), meaning the majority of the data are concentrated towards the lower average walking speeds. Furthermore, it shows a positive excess kurtosis for males on the single carriageway and females on a double carriageway. This indicates a higher unpredictability to measure walking speeds, since that compared to a normal distribution, there is a higher probability of extreme walking speeds. In addition to that, it shows a negative excess kurtosis for females on the single carriageway and males on a double carriageway. Compared to a normal distribution, this indicates a probability of walking speeds on the lower side (Cooper, 2022). However, one critical constraint presented above, is the difference between normal and crossing walking speeds being strictly categorised by gender.

2.3.3. Age Groups

Various research studies indicate a direct relation between age, perception, and cognition abilities on road-crash rates (Oxley et al. 1996; DeLucia et al., 2003; Helmers et al., 2004). Compared with the young adolescents, evidence proves that the elderly make slower traffic judgements (Oxley et al., 1996; Tung et al., 2008; Tournier et al., 2016; Feliciani et al., 2020). Alongside, data collected from Europe shows the elderly, especially for those aged above 80 years old, to form the largest group in pedestrian road-crash fatalities, ranging from 42% up to 52% (TSF, 2018). Similarly, in the UK, the number of pedestrian road-crash fatalities for those aged 60 and over scored 42% (DfT, 2015). The NHTSA has reported that the elderly have an increased risk of road-crash casualties. This is perhaps because of their increased fragility, slower movement, and declines in anticipating behaviour and perception

(NHTSA, 2015). Moreover, their death risk due to a road-crash at 60 km/h is high (ca., 90%) whereas adults and children present a surviving probability above 70% (Ashton and Mackay, 1979; Davis, 2001). By contrast, in Cyprus, pedestrian road-crash fatalities for pedestrians aged 65 and more scored the highest road-crash casualties in Europe, reaching 71% (Traffic Safety Fact Sheet, 2018). The second largest age group exposed to road-crash rates are children (Dellinger, 2002; Tay, 2003). According to Graham and Glaister (2003), child pedestrian road-crashes often occur in more deprived areas in which they appear to be less supervised by adults.

Table 2.3, illustrates some of the findings by age group of a study that was carried out in Izmir, Turkey, on pedestrian walking speeds[1].

Table 2.3.: Walking speeds by age group

Walking speed by Age Group	*Walking/15th percentile* (m/s)	*Crossing 15th percentile*
Children: 1-14 years old	Data not found	N/A
Young: 15-24 years old	1.36/1.10	1.61 1.33
Adults: 25-59 years old	Data not found	N/A
Elderly: 60+ years old	1.14/0.88	1.36 1.08

Source: *(Onelcin et al., 2017; Chandra and Bharti, 2013)*

As can be observed in Table 2.3, the results from the pedestrian walking speeds from six signalized intersections show, compared with the elderly the walking speed rates are higher for the younger when crossing the road (Onelcin et al., 2017). Detailed data on walking speeds involving children and adults were not found. Similar to Table 2.2, a key constraint of the above is the difference between normal crossing

[1] the percentile (15th) represents the value below which 15% of the scored walking speeds fall in. This means and 85% of the scored walking speeds are higher than the presented walking speeds (Javaid et al., 2022)

walking speeds, being strictly categorised by the age group. NHTSA has reported 45% of road-crash fatalities to non-vehicle-occupant involving children aged with up to 14 years old to happen outside of active traffic (i.e., parking lots, driveways, and private property) as a result of reversing vehicles (TSF, 2009). Generally, road-crashes take place on public roads. Yet, this study indicates partial responsibility on private property, where some of the non-reported cases to NHTSA, were concluded as possible accidents in private property (e.g., parking lot) (NHTSA, 2014, Noland et al., 2017). One example is, children's of short stature may limit their ability to see moving vehicles and impede their vision by parked cars or other obstacles (Ampofo-Boateng and Thomson, 1991; Rouse and Schwebel., 2019). Additionally, the small stature of children places them at greater risk for serious and fatal head injuries due to the alignment of car bumpers (Agran et al., 1989; Viano and Parenteau, 2021). In 2017, in Birmingham, USA, an observational study in the parking lots of six family/community recreational centers of 124 children aged between 2 and 10 years old, was carried out. The results showed that 67% of the child pedestrians aged between 2 and 10 years old were unsupervised (Rouse and Schwebel, 2019). Another study in New Jersey, USA, which has the highest share of pedestrian deaths, concluded that road-data results are unreliable and inconsistent. Therefore, the provision of preventive measures might apply (Graham and Glaister, 2002; De Oxley et al., 2005; Schuurman et al., 2009; Jones and Anciaes, 2017).

Age Groups: Young Adolescents vs. Elderly

This section refers to children and elderly pedestrians as the most vulnerable in a road environment. Through predictive simulation, the behavior of the young and the elderly pedestrians was contrasted (Weng et al., 2006; Feliciani et al., 2020). A simplified diagram is presented in Figure 2.3.

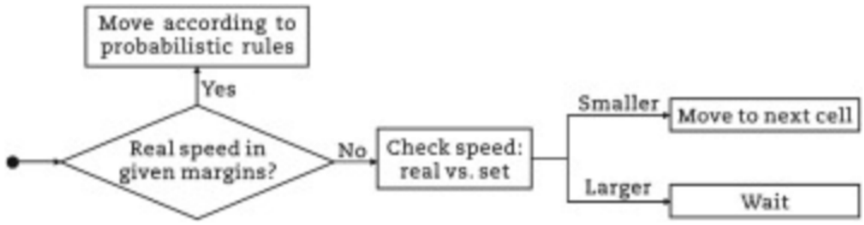


Figure 2.3.: Predictive simulator: young adolescents vs. elderly

Source: (Feliciani et al., 2020)

The illustrated diagram in Figure 2.3 represents the predictive simulation, which takes into account different pedestrian walking speeds and also allows for reducing walking speed fluctuations. First, the pedestrian movement compares probabilistic rules with real movement speeds with a pre-defined threshold value (Feliciani et al., 2017). The threshold value is defined according to probabilistic rules (Feliciani et al., 2017; (Feliciani et al., 2020). Additionally, it includes different reported crossing actions such as approaching, appraising and crossing. The pedestrian movement value below the threshold, activates the movement of the pedestrian. Otherwise, the pedestrian movement value above the threshold, forces the pedestrian to stop. However, with this predictive simulator, it not to be possible to accurately model quick changes regarding the walking speeds (Gorrini et al., 2018). Another disadvantage of this predictive simulator, was for not accounting the shortcuts often made by pedestrians when crossing the road (Feliciani et al., 2020). To support this type of special requirements, gathering field evidence have been previously recommended

(Feliciani et al., 2017). A more detailed examination on a number of previously applied pedestrian safety models by gender and age group are later presented in Chapter 4.

2.4. The Role of Road Infrastructure on Pedestrian Safety

Road infrastructure plays a vital role in pedestrian safety. GSRRS (2018) reported the UK with 0.665 pedestrian road-crash fatalities per 100,000 inhabitants in 2015. In the same report, in terms of road safety performance, Sweden was recognised as a world leader. In 2016, out of their road-crash fatalities per 100,000 inhabitants, 0.435 involved pedestrians. Moreover, between 1990 and 2015, the number of road traffic deaths in Sweden steeply dropped by 66% (STA, 2015). Comparably, in the UK, between 1991 and 2015, the number of road traffic deaths considerably dropped by 37,8 % (DfT, 2015). By contrast, Liberia, with 52.03 per 100,000 inhabitants, has been represented as one of the countries with the highest road-crash rates worldwide. Detailed road-crash data from Liberia with categorised road user locations (e.g., pedestrians, cyclists) was not found (GSRRS, 2018). With respect to infrastructure (e.g., road geometry, road condition), a study conducted in Sweden proved road infrastructure to reduce at least 59% of fatal road-crashes (Stigson et al., 2008). That said, generally pedestrian road-crashes occur when crossing the road, especially when pedestrian facilities are not made available (Bungum et al.,2005).

An illustration of the main types of road elements explored is illustrated in Figure 2.4.

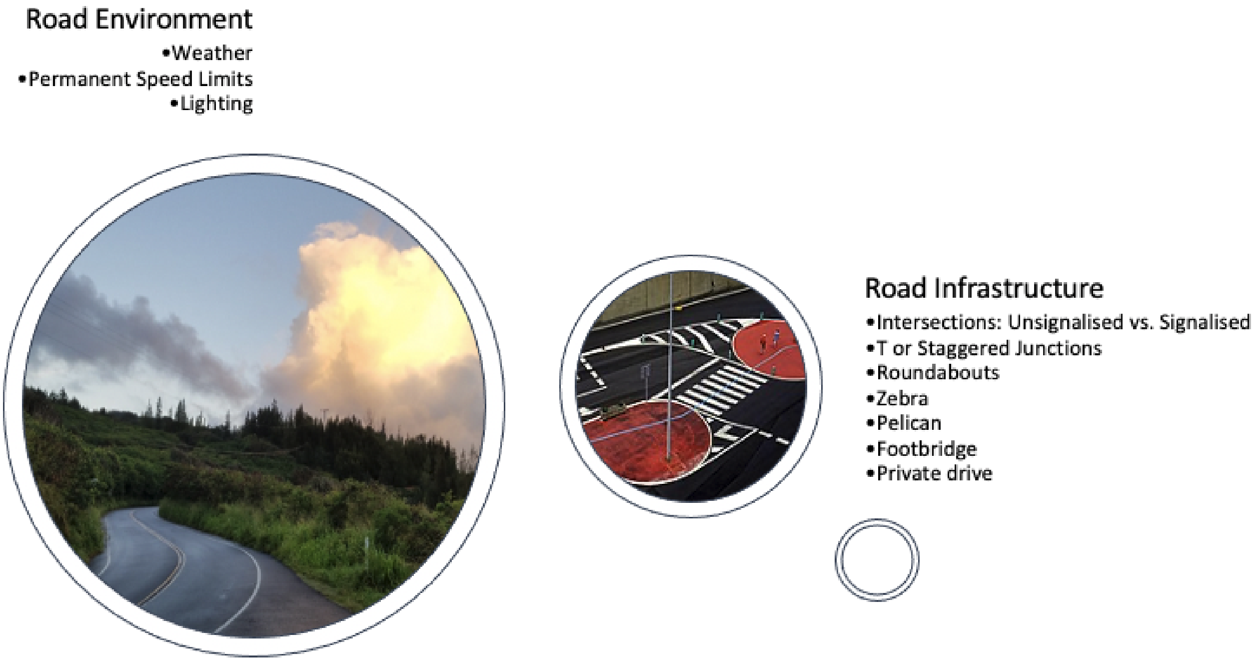


Figure 2.4.: Road elements

Figure 2.4 presents road features that, according to STATS19 include weather, speed limits, lighting an hazards (DfT, 2020). Regarding road infrastructure, Elvik et al. (2009) point out a number of infrastructural elements (e.g., geometrical design, pedestrian facilities) significantly affect pedestrian safety (Stigson et al., 2008, and Yannis and Karlaftis, 2010) [see Section 2.4]. This chapter includes exploring different types of road intersections (e.g., unsignalised, signalised) as well as different types of pedestrian physical facilities (e.g., zebra, pelican, private driveway) that may impact pedestrian safety.

2.4.1. Road Environment

Provisioning an appropriate road environment seems highly challenging, especially when attempting to anticipate pedestrian road-crash casualties (Theofilatos et al., 2017; Fotios and Gibbons, 2018; Malin et al. 2020). A more detailed description on some factors such as weather, speed limits and lighting condition, that likely affect pedestrian safety is presented next.

2.4.2. Weather

The weather condition has been previously recognised as a major road environmental factor contributing to pedestrian road-crashes (Kim et al., 2010). One of the reasons could be due to adverse weather conditions that cause road friction and visibility and therefore increase the risk of accidents (Theofilatos et al., 2017). Another reason could be that poor visibility and low pavement friction seem to increase pedestrian road-crash risk due to rainfall frequency and intensity (Mohamed et al., 2013; Theofilatos and Yannis, 2014). Moreover, side pavement ramps in the winter have been previously reported as hazardous for pedestrians (Mohamed et al., 2013; Theofilatos et al., 2017). According to Yannis and Karlaftis, (2010), in Athens, Greece, a study using 21 years of daily count data, demonstrated that increases in rainfall showed a decreasing tendency of road-crash fatalities. Interestingly, in 2017, the GB road data show most road-crashes with fine weather (84.3%) followed by raining (7.7%) (DfT, 2018). Chan et al. (2006) hold the view that physical activity levels can decrease in contrast to the increase in rain and snowfall (Togo et al., 2006). Turning into a study conducted by Ryser and Halseth (2019), in Prince George, British Columbia, Canada, the climate influence over the road design principles, for the city redevelopment in the winter season was explored. As a result, a review of 18 studies claims a lack of pedestrian accessibility due to poor design and undermined maintenance of pedestrian physical facilities (Owen et al., 2004; Ryser and Halseth, 2008). However, disappointingly, the pedestrian physical facilities accounting for weather conditions seem to remain limited and occasionally neglected (Wennberg et al., 2009; Giles-Corti et al., 2011; Klenk et al., 2011). Since it is almost certain that weather forecasting might not always be accurately assessed or reported, the provision of previously recommended multi-season pedestrian facilities (e.g., side pavement ramps to access sidewalks and puddles at street crossings) is likely applicable in the future (Li et al., 2013; Naik et al., 2016).

2.4.3. Speed Limits

Regarding speed limits, previous research studies have proven that decreasing speed limits can significantly reduce pedestrian road-crashes (Nilsson, 2004; Elvik et al., 2009). For instance, various research studies suggest permanent speed limits set at 40 km/h to be safer for the pedestrians (Elvik et al., 2009; Martin and Wu, 2018). Pasanen (1992) claims that compared with speed limits at 30 km/h, speed limits at 50 km/h may lead to up to eight times more pedestrian road-crash casualties. Furthermore, a more in-depth analysis regarding pedestrian street crossing speeds and awareness at an intersection, shows the significance of speed limits and proposed to define safety margins (Onelcin and Alver, 2017). However, from this study, it was not possible to accurately measure the behaviour of the pedestrians (Feliciani et al., 2017; Onelcin and Alver, 2017). GB road data from 2017 collected from STATS19, shows most road-crashes at 48.3 km/h (63%) followed by 96.9 km/h (10.1%). Likewise, between 2014 and 2017, in Finland, the results from a road-crash data-driven study highlighted pedestrian fatality rates mostly in zones with speed limits at 50km/h (22%), and at 70 km/h (35%) (Malin et al., 2020). In 2016, in São

Paulo, Brazil, a study to examine the introduction of variable speed limits at various signalised and non-signalised intersections, was carried out. Interestingly, the results show that reducing speed limits at signalised intersections influenced the residents' route-travel choices, which were noticed to diverge from treated intersections to alternative routes (Ang et al., 2020). A fundamental issue here, can be delimiting and defining boundaries for implementing traffic calming treatments at the most relevant locations (Granà et al., 2010; Bichicchi et al., 2017). Therefore, advancing alternative systems to help improve the implementation of traffic calming measures more accurately has been previously proposed (Nilsson, 2004; Kangas et al., 2015).

2.4.4. Lighting

There is an imminent need to further investigate the effects of lighting on pedestrian road-crash in an urban road environment (Sullivan and Flannagan, 2002). Generally, road lighting systems support pedestrians to avoid accidents through enabling them noticing hazards. Also, to make it possible for them to recognise other road users while giving them a sense of security, especially at night where pedestrians were found up to seven times more exposed to danger than during daytime (Fotios and Castleton, 2016; Fotios and Gibbons, 2018). According to Bently and Haslam (2001), road lighting is an essential aid to detecting obstacles on the side pavement in public roads. This is because it might result in reducing falls, which is significant in terms of the severity of the cases resulting in fatalities. The lighting level on the roads, are in general described on a range of 5 possible locations where an accident occurred and led to a pedestrian fatality (Fotios and Gibbons, 2016; Fotios and Gibbons, 2018). A study to measure the ability to detect peripheral obstacles under variations of light (i.e., illuminance), for example the illuminance conditions ratio between the vision of and eye under nightlight (i.e., scotopic) and the eye under normal daylight (photopic) was carried out (Fotios et al., 2016). The results show that, higher illuminance ratios increase the probability of detecting obstacles. Moreover, the lighting for other visual needs of pedestrians to reassure their cognitive judgements proved to be necessary (Uttley et al., 2015; Fotios et al., 2016). One example is that, from 1996 to 2008, the results from a study carried out in Greece show that from 358,485 road-crashes recorded by the police, fewer pedestrian road-crashes occurred on lit roads than on unlit roads during the night, proving that lit roads at night have a greater potential for reducing pedestrian road-crash fatalities (Yannis et al., 2013). Another study over a 4-year period (1999–2002) in Minnesota, USA, shows 22,058 road crashes at 6464 intersections. Additionally, the results specified that both at night, and during the day, illuminated (i.e., lit) intersections share 12% of less road-crash rates than non-illuminated (i.e., unlit) intersections (Bullough et al., 2013). In 2017, the UK road data showed most road-crashes at daylight (44.2%) followed by at night lit (36.6.%) and night unlit (14.9%) (DfT, 2018). However, disappointingly there was also a significant amount of unknown illumination conditions at night (ca. 25.5%).

2.4.5. All Roads

Evidence from previous studies shows pedestrian road-crashes to happen more often when pedestrians are located at signalised intersections in urban areas (Feliciani et al., 2017; Onelcin and Alver, 2017; Gorrini et al., 2018). Additionally, Keegan and O'Mahony (2003) reported that 35% of the pedestrians illegally entered a signalised crossing. A study on pedestrian road-crashes at crossing facilities in New South Wales and Victoria, Australia, found that illegal pedestrian movements featured in 32–44% of pedestrian crashes at signalised intersections and 45% at pedestrian operated signals (i.e., not at a signalised intersection) (Senserrick et al., 2012). Over an 11-year period (1996–2006) between 8 a.m. and 6 p.m. on weekdays, at intersections and on the roads approaching them: 77 crashes (41.8%) were recorded when the pedestrians were crossing legally; 64 crashes (34.8%) when the pedestrian crossed within 20 m of the signalised crossing; and 43 crashes (23.4%) when the pedestrian entered the crossing traffic light zone at either the intermittent (i.e., amber) or steady (i.e., red) traffic light. However, over 58% of illegal pedestrian movements were not recorded (Austroads, 2018). In 2017, UK road-crashes ranked the highest scores at intersections, particularly in urban environments (DfT, 2017). Additionally, the STATS19 road data show most road-crashes occurring when the physical infrastructure is not made available (54.5%), followed by at junctions (e.g., T, staggered intersections) (23.7 %) (DfT, 2017). By contrast, the roundabouts represented the lowest road-crash rates (Wisch et al., 2019; Wu et al., 2019; DfT, 2019). More specifically, Turner et al., 2012 and Lluri and Golgota (2019), draw attention to the potential of replacing a three-arm or four-arm intersection by a roundabout, leading to a reduction of up to 30% of road-crashes. This is because, roundabouts seem to provide more satisfactory pedestrian safety levels when operating at low traffic volumes and speeds. Another reason is that, evidence from an automatic emergency breaking testing system created in Europe, identified 18% of the fatalities at a junction (e.g., T staggered) in number of European cities (Wisch et al., 2019). These findings suggest the advancement of road infrastructure, by developing a series of pedestrian road-crash models for different types of intersections (Lee et al., 2005).

Road Intersections

Most countries in Europe differentiate the road intersection layouts. However, in terms of geometry and layout, their comparability seems limited (Wisch et al., 2019). Figure 2.5 illustrates different types of road intersection layouts under analysis.

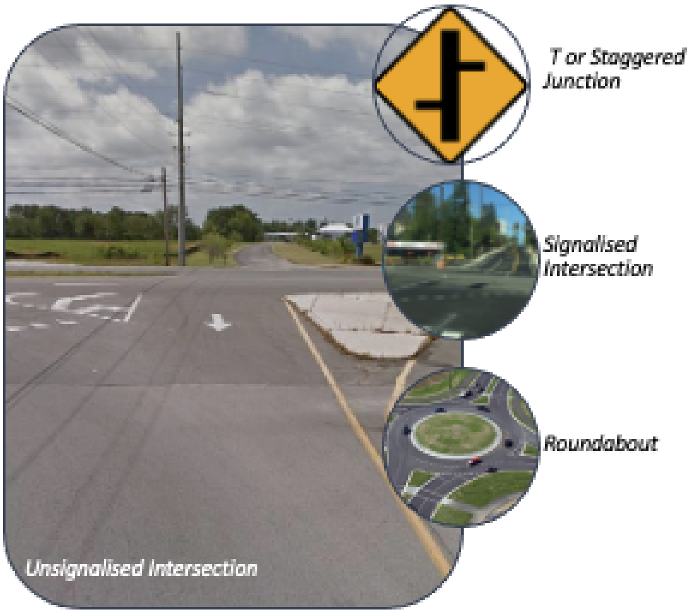


Figure 2.5.: Road intersection types

The road intersections illustrated in Figure 2.5 are described as unsignalised, signalised, staggered junctions, and roundabouts. According to Prasetijo and Ahmad (2012), unsignalised intersections are designed to regulate a low volume of traffic flow in an heterogeneous traffic condition. Often, they are used as a popular method to help reduce dangerous *jaywalk*, in cases where urban physical infrastructure is not provided. Both the low costs of their installation and the technical requirements, turn them into a flexible solution to improve pedestrian mobility in urban areas (Gorrini et al., 2018). Compared with unsignalised intersections, signalised intersections are more expensive as they make use of traffic lights with three light crossing phases for pedestrians (Feliciani et al., 2017; Onelcin and Alver, 2017). Still, they are usually preferred particularly in dangerous areas (e.g., low visibility) or in high traffic volume zones (Feliciani et al., 2020). One key limitation, is the underestimation of the pedestrians' crossing speeds within the given time. In Izmir, Turkey, data from a study conducted in 2013 at six signalised intersections revealed the walking speeds of 334 pedestrians when crossing a 25-meter intersection. Interestingly, the age did not reflect a significant effect on the walking speeds (Onelcin et al., 2017). A key limitation was anticipating pedestrian crossings (e.g., walking speeds due to a level impairment). This is critical when attempting to more objectively implement pedestrian safety measures [see Sections 2.3, 2.3.1]. Regarding T or staggered junctions, according to Li et al. (2012) these have represented the highest road-crash levels, due to poor sight distance conditions on multiple conflict points. Also, in 2017, GB pedestrian road-crashes scored the highest rates at intersections particularly in urban road environments (DfT, 2019). In contrast, with respect to roundabouts, in various countries and frequently, these have demonstrated significant safety advantages over other types of intersections (e.g., three-arm or four-arm non-signalised junctions). O'Cinneide and Troutbek, (1998), Hydén and Várhelyi (2000), and also Vijayawargiya and Rokade (2017), point out that roundabouts and mini-roundabouts used in urban areas can reduce conflict points between traffic flows and moderate driving speeds (Vignali et al., 2020). The authors also claim that, replacing a three- or four-arm intersection by a roundabout seems to lead to a reduction of up to 30% of pedestrian road crashes. However, disappointingly studies related to roundabout safety have generally overlooked the pedestrians safety by giving more focus to the drivers (Cohen et al., 2013; Gross et al., 2013). Nevertheless, installing guardrails across roundabouts as a further countermeasure has been previously suggested to achieve a greater efficiency (Gross et al., 2013; Vijayawargiya and Rokade 2017).

Pedestrian Physical Facilities

Recent research studies have identified the absence of pedestrian physical facilities as a major factor impacting on pedestrian safety, especially at illegal street crossings (Bungum et al., 2005; Bauer et al., 2016). Pedestrian physical facilities, are developed so that road users (e.g., drivers, cyclists, pedestrians) can negotiate a shared space (e.g., change a lane, cross a road) within a road environment (Lluri and Golgota 2019; Wisch et al., 2019). However, pedestrians have been found crossing pedestrian facilities at the red traffic light (Li et al., 2005). One advancement in pedestrian facilities is the concept of shared spaces based on greater segregation between the facilities for different types of road users, has been developed as a more pedestrian-friendly road infrastructure than conventional street layouts (Kaparias et al. 2012; Lee and Kim, 2019). Furthermore, recent advancements on forgiving roads have remarkably improved pedestrian safety (Cohen et al., 2013; Gross et al., 2013; Godavarthy and Russell, 2016; Vijayawargiya and Rokade, 2017). Forgiving roads are essentially designed to offer resilience to human errors so that the consequences of traffic accidents are mitigated (O'Hern et al., 2019; Tsigdinos et al., 2022). One example is by taking into account crossing and walking speeds, as demonstrated in Section 2.4.5. This has been confirmed by Sarkar et al. (2011), Sasidharan et al. (2014) and Wu et al. (2019). To support advancements of forgiving roads, in addition to infrastructural measures (Bichicchi et al. 2017; Malkhamah et al., 2005; Oviedo-Trespalacios et al., 2017; Costa et al., 2018; iRAP, 2021), data-driven initiatives have been previously proposed (Malkhamah et al., 2005; Cohen et al., 2013; Gross et al., 2013; Godavarthy and Russell, 2016; Vijayawargiya and Rokade 2017; DfT, 2017). In addition to that, combining the use of different types of technologies over the road infrastructure has also been proposed (Godavarthy and Russell, 2016). Interestingly, to ensure a more systematic approach, complementary methods have also been suggested (Cahill, 2010; Wisdom and Creswell, 2013). As an example, the adoption of preventive measures such as educational programs and outreach activities have been put forward (Tay, 2003; McMillan et al., 2005; Urie et al., 2016; Lin et al., 2019). Turning into GB, road data show most road-crashes happening when pedestrian physical facilities were absent (67.6%), followed by pedestrians located at a pelican or similar (14.1%) (DfT, 2018). Figure 2.6 illustrates different types of pedestrian physical facilities that were previously implemented (iRAP, 2021; DfT, 2020).

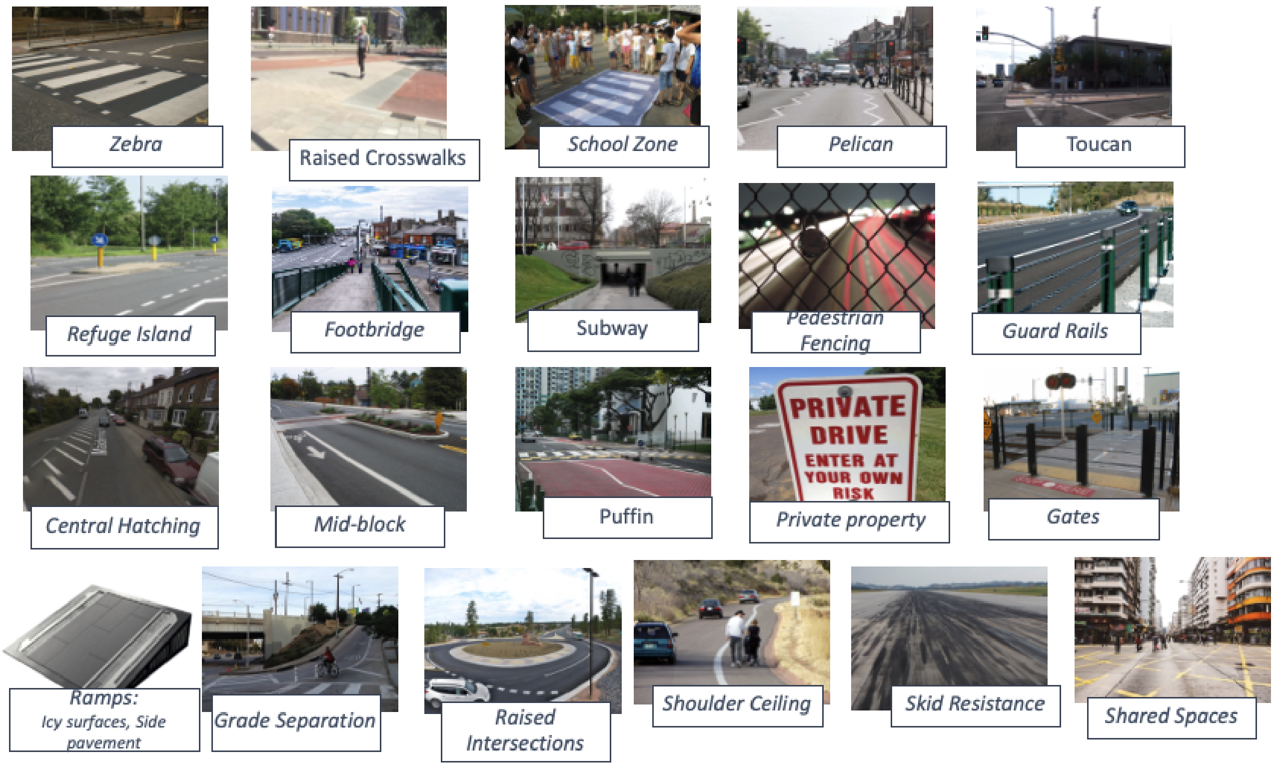


Figure 2.6.: Pedestrian physical facilities

The types of pedestrian physical facilities further investigated in this study are the following: zebra, pedestrian light controlled crossing (pelican), refuge, subway or footbridge, and private property.

i. Zebra: used to give priority to pedestrians. In this case, drivers are obliged to stop when someone has indicated their intent to cross by waiting by the crossing.

ii. Pelican: where the flow of traffic is controlled by lights.

iii. Central refuge: typically used when a street is very wide. It is a small section of pavement, surrounded by asphalt or other road materials, where pedestrians can stop before finishing crossing a road.

iv. Footbridge or subway: these two types of physical facilities are designed to allow pedestrians to reach the other side of the road in safety. The footbridge is a bridge designed solely for pedestrians whereas the subway is and underpass for pedestrians beneath a road.

v. Private driveway: private road for local access to one or a small group of people. Usually owned and maintained by an individual or a small group of people.

The different types of pedestrian physical facilities presented above, have been previously explored in Section 2.4.5. According to Bichicchi et al. (2017), zebra crossings are the best perceived elements by all the drivers. One example is data collected from several studies across different Indian cities, showing pedestrian road-crash casualties ranking up to 40%, particularly where pedestrian physical facilities (e.g., zebra, footway) were not present. Moreover, it also indicated the presence of zebra crossings, reducing road-crashes remarkably (Huang et al., 2000; Mohan et al., 2016). However, on the contrary, in Singapore 22% of all pedestrian fatal accidents occurred at signalised and designated pedestrian crossing facilities, which seems rather intricate (Koh et al., 2014). Chen et al. (2009) point out that high volumes of pedestrian crossing also increase the pedestrian road-crash risk. Regarding the pelican facilities, these are mostly used to extend the clearance period while pedestrians are still on the crossing. Moreover, Malkhamah et al. (2005) propose the adoption of pelican crossings for implementing automated pedestrian road-crash detection. A key limitation of the pelican is its inability to capture the behaviour of pedestrians when involved in a road-crash. Concerning the central refuge, Zegeer and Bushel (2012) claimed island or median refuge to reduce pedestrian road-crashes by circa 18%. However, according to Costa et al. (2018) and Yuan et al. (2011), the island refuge shows limitations in terms of visibility. The authors proposed vertical signs in front of the driver to ease their focus while driving. In Barranquilla, Colombia, between 2014 and 2015, as a support area to cross over a highway, a study with 210 participants on a footbridge was carried out (Oviedo-Trespalacios et al., 2017). Even though the results show most of the participants are aware of their increased safety when crossing the footbridge (instead of the highway), only a third of the participants responded to make use of the footbridge to cross the highway. One possible reason could be related to the exposure to crime and violence (Oviedo-Trespalacios et al., 2017). With respect to private property (e.g., parking lots, driveway), in New Jersey, the USA, the results from a study on pedestrian road-crash casualty in private property were concluded as unreliable and inconsistent. Still, these indicate pedestrian road-crash mostly related with children (Noland et al., 2017) [see Section 2.3.3].

2.5. Learning Classification Algorithms applied to Road Data

The datamining classification methods examined in this study, distinguish two main types of data analysis: continuous and discrete. Generally, the continuous data analysis includes any value from a given range (without any gap) (Zeger and Liang, 1986; Shirkhorshidi et al., 2015). A key disadvantage, is when selecting the most applicable variables to carry out an analysis from a large variety of similar measures, it may lead to confusion. Additionally, the metrics may perform differently for datasets with diverse features (e.g., pedestrian behaviour, pedestrian physical facilities) (Shirkhorshidi et al., 2015). Regarding the discrete data analysis, in principle it makes use of a certain number of isolated values (Fiorentini and Losa, 2020). To improve its accuracy, the exploitation of a wider range of different algorithms as well as a more sophisticated calibration of the algorithms, have been previously proposed (Max and Jonhson, 2013; Wang et al., 2019; Zheng et al., 2019). As the data for this study refers to isolated values from distinct data sources, distinct types of discrete ML algorithms will be further explored and described.

2.5.1. Supervised-learning Algorithms

Supervised-learning algorithms are used to classify data variables (e.g., pedestrian attributes, pedestrian facilities). Also, labeled instances (e.g., gender = 0, 1; road-crash fatality severity = 0, 1, 2) can be obtained. However, typically this type of model demonstrates high dependency on historical events as well as data quality (Soilán et al., 2018). A key limitation of this method is its inability to accurately differentiate road users (e.g., pedestrians and cyclists). However, depending on the type of road user, the overall accuracy rates (ca. 88%) were noticed to differ: 80% for pedestrians and 60.7% for cyclists (Zangenehpour et al., 2015). Another challenge is the issue of converting the findings into information that can directly assist decision-making undertaking relevant pedestrian safety measures (Serna et al., 2014; Soilán et al., 2018) even when being study specific (i.e., applicable to a specific location).

2.5.2. Unsupervised-learning Algorithms

Unsupervised-learning algorithms are used in such a way that each class element corresponds to a cluster (e.g., pedestrian age group). Each cluster is interpreted according to the categories of the most relevant variables (Ballesteros et al., 2004; Kim et al., 2008; Prato et al., 2012). The greater advantage of implementing unsupervised over supervised-learning algorithms is its ability to delineate patterns independently with a lower dependency on past data as well as the number of examined variables (Cottrell and Rousset 1997; Prato et al., 2012). Additionally, in the context of road crashes, it accounts for different risks related to different regions based on regional use and activity patterns (Elias et al., 2010). Still, as a complementary activity, conducting qualitative research (e.g., audits, surveys) with a particular focus on road infrastructural countermeasures (e.g., narrow roads) has been previously proposed (Cahill, 2010; Prato et al., 2012; Wisdom and Creswell, 2013). A more detailed examination on a number of previously applied pedestrian safety models by weather, speed limits, lighting, road intersections and pedestrian physical facilities, are expanded in Chapter 5.

2.5.3. ML Algorithms: Supervised vs. Unsupervised

The described ML algorithms (i.e., supervised, unsupervised), present the supervised-learning ML technique with a high dependency on historical events (Soilan et al., 2018; Sanz-Casado et al., 2019; Assi et al., 2020). By contrast, the unsupervised-learning ML technique displays a lower dependency from past data. In addition to that, it enables to account different risks related to distinct regions based on local patterns (Elias et al., 2010; Prato et al., 2012; Feliciani et al., 2020; Giummarra et al., 2021). Since the present study was designed to adopt ML to examine new patterns using a lower number of historical data (i.e., data from 2017), the unsupervised-learning ML may be the most appropriate to conduct the proposed analysis. However, another important finding is that, for both supervised and unsupervised algorithms, the lack of data quality seems to remain crucial to enable replicating real world environments (Imprialou and Quddus, 2019). Nevertheless, prior studies have noted the importance of pedestrian safety modelling (PRACT, 2016). However, less was found in the literature on the question of adopting models that can jointly examine multiple risk factors leading to road crash-casualties. This means that to date each predictive model (e.g., Negative Binomial Regression; Integer Autoregressive Model) addresses a single risk factor separately (Yannis and Karlaftis, 2010; Zegeer et al., 2020). This will be later on shown in Chapters 4 and 5. On the contrary, in this study unsupervised-learning will be explored to simultaneously allow for examining multiple risk factors (Feliciani et al., 2020; Giummarra et al., 2021). This will be demonstrated in Chapter 7.

2.5.4. Road Attributes Definition: Data Collection

Advancements on examining data impacting on pedestrian safety (i.e., pedestrian attributes and road elements), have been previously performed (Jagnoor and Peden, 2020; Earl et al., 2018; Oxley et al., 2005; Nemire et al., 2016; Cahill, 2010; Papadimitriou et al., 2015; Wisdom and Creswell 2013). Still, to provide relevant recommendations, data quality, validity and applicability, is fundamental (Bungum et al., 2005; Oxley et al., 2005; Cahill 2010; Saunier, 2013; Nemire et al., 2016; Earl et al., 2018). This study focuses on examining road data as a preventive measure (i.e., pre-crash)(Haddon, 1970; DfT, 2019). The STATS19 database has been advanced by the Department for Transport (DfT) in the UK, and it is made available to provide an overall measure of road safety. The data can be used to help identify traffic safety problems and to provide an objective basis to, evaluate the effectiveness of pedestrian safety measures while improving road safety programmes (DfT, 2019). Therefore, to examine road fatality records involving pedestrians, three main categories were investigated more in depth: pedestrian general attributes (e.g., gender, age group), the road environment (e.g., lighting, speed limits), and the road infrastructure (e.g., road type, pedestrian physical facilities) (Sections 2.3 2.4). Moreover, one crucial aspect regarded exploring data with a model is a lower level of dependency on historical events and data quality [see Sections 2.2.1 and 2.2.3]. The methodology will be explained in Chapter 3, and its applicability will be further explored in Chapters 6 and 7, experimented in Chapter 8 and further tested in Chapter 9.

2.6. Summary and Identification of the Knowledge Gaps

This chapter provides insights into the effects of incorporating discrete learning classification algorithms to systematically examine pedestrian road-crash casualties. This new study adopted data-learning algorithms that provide a pioneering examination of pedestrian attributes and road elements that affect pedestrian safety. Since to date, the reasons for the occurrence of pedestrian road-crash casualties seem to not be explicitly clarified, this project provided an important opportunity to advance the understanding of a fundamental aspect to serve pedestrians in a road environment. Returning briefly to the pedestrian general attributes, while commuting, the level of impairment of the pedestrians may directly influence their awareness of their own vulnerability to be able to protect themselves appropriately (GRSP, 2008) [see Section 2.3]. Mayou and Bryant (2003) drew attention to the noted causes of the diversity of road fatalities, due to among others, psychological factors across various vulnerable groups. Therefore, it was felt important to exemplify the impact of a data-driven predictive simulation contrasting the movement of young and elderly pedestrians (Bohannon, 1977; Tay, 2003). One drawback of the simulator was that it failed to anticipate shortcut routes likely taken by pedestrians when crossing the road [see Section 2.3.3]. Turning into the road elements, recommendations to explore both environmental (e.g., weather, climate, speed limits) and infrastructural (e.g., geometric, traffic, physical facilities), have been previously suggested (Kim et al., 2008; Obeng and Rokonuzzaman, 2013; Shaheed et al., 2013; Fotios and Gibbons, 2016; Fotios and Gibbons, 2018). Turning into data-learning algorithms, these have proven their ability to examine pedestrian road-crashes more efficiently (Ballesteros et al., 2004; Kim et al., 2008; Elias et al., 2010; Prato et al., 2012; Li et al., 2020). Moreover, provisioning modelling (e.g., ML) road design (e.g., geometry) and implementation (e.g., lighting) have demonstrated their vital capabilities to improve pedestrian safety (Bullough et al., 2013; Fotios and Gibbons, 2016; Fotios and Gibbons, 2018). Regarding modelling, existing measures could be enhanced through adopting more sophisticated methods (Nilsson, 2010; Fotios and Gibbons, 2016; Fotios and Gibbons, 2018). Furthermore, unsupervised data-learning algorithms require a lower dependency both on human intervention and on historical data, they have the potential to expand their applicability to other regions based on local patterns. This research adopts unsupervised data-learning algorithms to systematically analyse GB pedestrian road-crash casualty causation from STATS19 (DfT, 2017; Imprialou and Quddus, 2019; DfT, 2018; DfT, 2019). Moreover, rather than examining historical data, this study seeks to test the applicability of ML on spatial pedestrian road-crash data which will help to analyse data from two specific regions (e.g., GB, and London), with a view to explore road-crash casualties from different regions, with a similar approach (Zhaid et al., 2020; Infante et al., 2023). This will be demonstrated later in Chapters 8, and 9. Together, these findings may contribute to a better understanding of the importance of advancing preventive measures for pedestrian safety that have not been considered to date. However, this study is unable to encompass all the elements affecting pedestrian safety. An example is the inability to measure the levels of cognitive impairment and physical disability. Another reason could be the failure to account for adverse weather conditions in conjunction with the absence of pedestrian facilities. On the latter note, the gathered evidence suggests that pedestrian infrastructure plays a crucial role in ensuring pedestrian safety.

3. Methodology

3.1. Introduction

The following sections will elaborate on the proposed methodology. Section 3.1 introduces the overall approach of the chapter. Section 3.2 introduces the proposed ML method to examine road data. It includes defining pedestrian attributes and road infrastructure elements, examining the actual data (e.g., attributes), and testing ML algorithms applicability and sensitivity response. Section 3.3 presents the ML approach including the data analysis process that was carried out. Section 3.4 describes the procedure to evaluate the adopted prototype to examine road data. Section 3.5 discusses the findings.

3.1.1. Overall Approach

There is a growing body of literature that recognises the contribution of data-driven analysis to prevent pedestrian road-crashes in urban road environments (Pilkington and Kinr, 2005; Deb et al., 2017) [see Chapters 2, 4, and 5]. Still, reasons underpinning pedestrian road-crash casualty causation are still unclear (Adminaite et al., 2015; Papadimitriou et al., 2019; Twisk, et al. 2015; Zhang et al., 2019; DfT, 2019;

OECD, 2020; WHO, 2020) [see Section 1.2]. Furthermore, to date, no single approach has clearly demonstrated direct implications from influential factors such as pedestrian attributes and road elements. One possibility could be the data inconsistency restraining the discovery of co-relational links among the investigated factors. This has been acknowledged as a critical disadvantage to validating any type of data-driven approach, particularly when attempting to expand new knowledge from purely data-driven sources (Nashad et al., 2016; Hariyono, 2017; Feliciani, 2020; DfT, 2020). Recent evidence suggests that ML has demonstrated its capability to supersede traditional data analysis techniques (e.g., forecasting) to examine road data (Sasidharan et al., 2014; Gao, 2017) [see Chapter 2]. One greater advantage of ML over conventional analysis is that, regardless of its nature, type or background, various studies have proven existing ML algorithms to more objectively identify patterns (e.g., data science, biological science, medical science, meteorology) (Trueck and Rachev, 2009; Bonga et al., 2020; Elaziz et al., 2020; Stirnberg et al., 2020). Another advantage shown from previous studies, is that when investigating road data, compared with other models such as the K-Nearest Neighbour (KNN) with an accuracy of up to 97.3%, ML proved to have the ability to reach up to 99% of accuracy. Additionally, when compared with traditional methods, for large data samples a higher accuracy by 1.8% was reached (Kaplan and Prato, 2012; Zhaid et al., 2020). However, it is important to note that the data interpretation in this study will be study specific (i.e., GB pedestrian road-crash casualty data). The chosen approach has considered two main aspects: road data availability from STATS19, and ML road data analysis. With respect to data availability, the Department for Transport (DfT) collects road data to produce statistics on the level of traffic (e.g., peak-rush hours) and accidents (e.g., non/serious injuries) (DfT, 2019). On the question of using ML to analyse road data, ML algorithms have proven their potential to delineate road-crash trends more efficiently. The first research question of this study regards defining features that influence road crashes. The adopted method independently examines data about pedestrians and road elements (Sasidharan et al., 2014; Wu et al., 2019; Cohen and Dalyot, 2020). In addition to that, to establish ML algorithms applicability in this study, the following procedure was carried out:

i. Defining elements (e.g., pedestrian attributes, road elements) that seem to influence the safety of the pedestrians based on the literature review findings.

ii. Extracting data from public databases (e.g., STATS19) followed by cleansing and analysing the respective data.

iii. Data learning, data modelling, data verification and data validation.

iv. Data interpreting, and data contrasting with the findings with different evaluation metrics from the literature review.

The literature suggests adopting ML techniques to improve pedestrian safety (Feliciani et al., 2020; Giummarra et al., 2021) [see Section 2.2]. However, the applicability of previously experimented ML methods may no longer be valid due to the uncertainty of the future, as for example, technology advancements (Trueck and Rachev, 2009; Noland et al., 2017; Ryserand Halseth, 2019; Zhaid et al., 2020) [see Section 2.6]. Two other major concerns are data inconsistency and the model validation process. Regarding the data inconsistency, lack of credible information from public data sources and traditional data extraction methods is expected. One example is the human intervention (e.g., a police officer recording pedestrian road-crashes incorrectly if distracted). About the model validation process, the data processing (e.g., cleansing) and the model assessment are two equally important elements, since their erroneous results (e.g., overestimated, underestimated) might disqualify the validity of the outcome findings. The newly proposed ML method has been chosen for the following two key reasons:

i. The method seems to overcome existing limitations of data sources in terms of data reliability and robustness (e.g., STATS19).

ii. Accordingly, this enables the method to anticipate patterns with minimal human intervention from both small- and large-scale datasets.

Compared with conventional data analysis, ML may supersede in terms of accuracy. Therefore, the proposal of this study is to incorporate ML models and investigate, in a specified period, factors affecting pedestrian safety (i.e., pedestrian attributes and road elements) in a road environment.

3.2. Methodology

Figure 3.1 presents the adopted methodology to conduct this study.

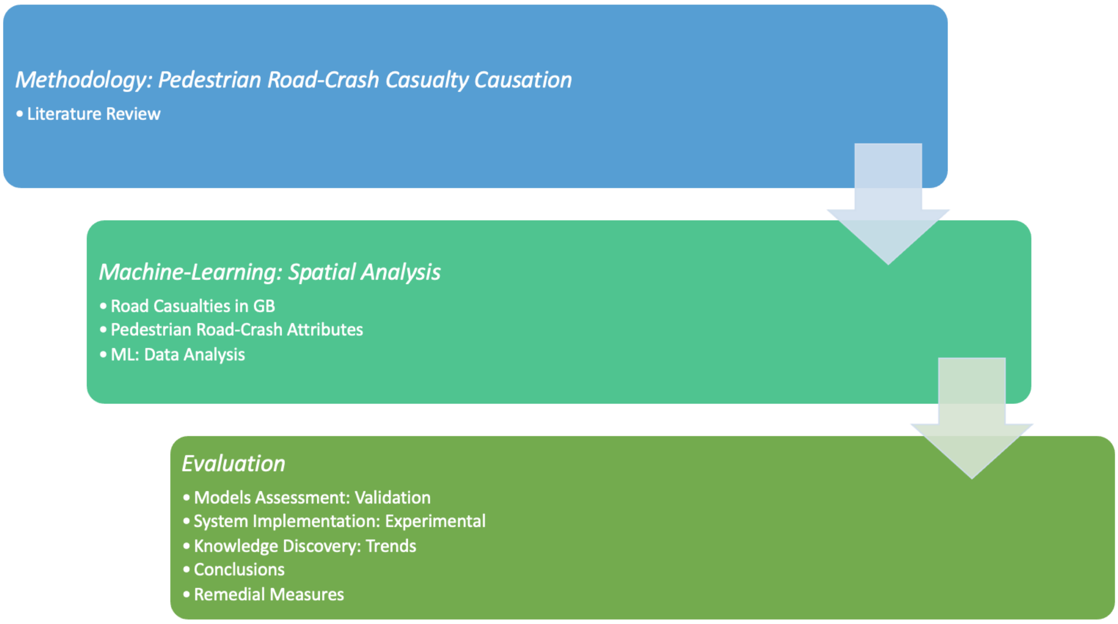


Figure 3.1.: Adopted methodology

Figure 3.1 displays the adopted methodology, which is divided into three main stages: background work on road-crash data, ML spatial data analysis, and evaluation. The literature review was carried out with regard to examining road-crash casualty trends. Additionally, it included a thorough investigation of a range of existing pedestrian safety ML models. The findings indicated potential influences of the different models or features adopted. The ML data analysis included data gathering, assessment, and verification. The ML data analysis defined pertinent elements (e.g., pedestrian attributes, road elements) to be analysed. To achieve that, road data from public data sources (e.g., STATS19) was collected. Thereafter, the selected pre-defined elements, later on described in Chapters 4 and 5, were converted into a computational format (i.e., data feed) (KNIME, 2021) [see Chapter 7]. The pre-defined elements are from now on referred to as attributes. These attributes were used to identify relationships between the pedestrian general attributes, the road environment, and the road infrastructure. Once the attributes were gathered, it was possible to perform the ML data analysis. Lastly, the evaluation stage was performed to compare the findings from the literature review with the findings of the experiments and validate these with expert opinion to conclude trends. It consisted of three functions: model assessment, knowledge discovery, and conclusions. Regarding the model assessment, their accuracy was validated and verified using different metric measures, such as Spearman's rho and Receiver Operating Characteristic (ROC) (Kwon et al., 2015; De Winter et al., 2016; Silva et al., 2020). These will be explained later on in Section 6.3. In addition to that, the introduction of new features through the adoption of strict co-relational rules applied to the road data used in this study was proposed. Regarding knowledge discovery, after contrasting the newly discovered knowledge with the literature review, the trends were rationalised. Thereafter, the relevant findings of this exercise are expected to be incorporated as a remedial measure or used as a recommendation (e.g., new feature, road management improvement).

3.2.1. Data Analysis

As presented in Section 2.2.2, a data analysis comprises four stages: data input, data processing, reporting, and data output (Noland et al., 2017; DfT, 2019). In this section, each of the stages of the spatial data analysis will be explained.

3.2.2. Spatial Analysis

A spatial analysis using a computational data tool was carried out. The examination entailed applying ML modules to one year of data (i.e., 2017). The findings were then compared with actual data from the year ahead (i.e., 2018) (Hiebl et al., 2016; FSV, 2019a). In terms of the ML algorithm accuracy rates (i.e., 97.03%), the results from the undertaken examination (i.e., spatial data analysis) surpass the minimum standard expectations (i.e., 70%) [see Section 8.3]. Additionally, for each of the examined factors, the undertaken data analysis establishes three distinct rules to test the model response, respectively, at 80%, 75%, and 85% [see Section 9.3]. To examine the accuracy of the system more in detail, the data from 2017 was evaluated spatially on a sample of two regions (e.g., GB, and London) (FSV, 2019a). This included examining in more detail GB road-crash data specifically for one year (e.g., 2017) [see Chapters 2, 8, and 9]. Furthermore, to conclude more coherent results, a traditional analysis to examine the data temporally (i.e., 2009–2019) was carried out. This is for homogeneity and to not compromise consistency when discussing the findings (Kann et al., 2012; Hiebl et al., 2016). For example, the findings of the 10-year analysis point to 2017 as the year with the highest increase rate (5%) of pedestrian road-crash casualties [see Chapter 10]. This is the reason why 2017 was chosen as the year to carry out the spatial analysis. It is important to highlight that the application of ML modules to compare findings with actual data from the subsequent 10 years (i.e., 2020–2030) is beyond the scope of this study.

3.3. Machine-Learning Approach

The adopted ML approach in this study explores new knowledge using two ML techniques: supervised-learning, unsupervised-learning. The discovered knowledge will be compared with actual data extracted both from traditional statistical methods and findings from previous work (DfT, 2017; DfT, 2018). The prototype using GB road data will be expanded in Chapter 8. In addition to that, the sensitivity response of the prototype will be analysed in Chapter 9 and validated with expert opinion in Chapter 10. Figure 3.2 displays the proposed ML approach to analyse pedestrian road-crash casualty data in GB [see Section 2.2].

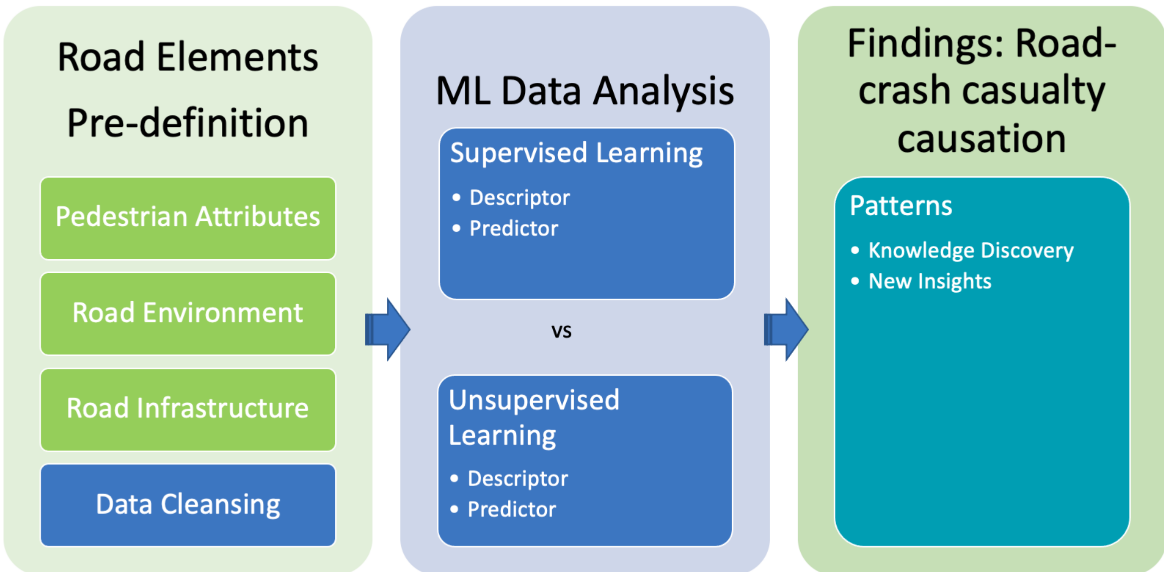


Figure 3.2.: ML data analysis: supervised vs. unsupervised

The ML approach adopted in this study illustrated in Figure 3.2, instead of extrapolating data through forecasting formulas from historical data, it incorporates ML algorithms to discover unforeseen patterns through inserting data (i.e., attributes, data feed) into a computational datamining tool. Here, aspects of importance are pre-defining and selecting elements by using rules and performing a data cleansing process (e.g., data filter) [see Section 2.2]. This will be explained later in this chapter. Once that is achieved, the ML data analysis can be introduced. This technique adds two key details: ML algorithms and data verification. About the ML algorithms, two different types of ML techniques were described. These are presented in Section 2.2.3. In addition to that, regarding the data verification, it applies two sub-modes: descriptor, and predictor. The descriptor trains the data to be fed to the predictor so that it can validate the data fed to project results (Assi et al., 2020). To understand how these sub-modes might influence the data, a series of comparative simulations were carried out [see Chapters 8 and 9]. Nevertheless, despite all the efforts, a key restriction of this approach might still be the choice of the most pertinent ML algorithms to best interpret the available data. The proposed methodology incorporating the ML data analysis is presented in the next section.

3.3.1. ML Data Analysis Process

The proposed ML data analysis process is illustrated in Figure 3.3.

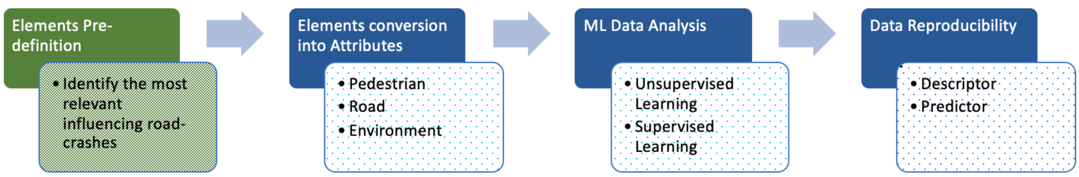


Figure 3.3.: Proposed ML process

Figure 3.3 illustrates the key stages carried out to conduct this study. These are best treated under the following 4 headings:

i. Pre-defined elements: a more in-depth explanation can be found in Section 2.2.

ii. Attributes: the process of converting the pre-defined elements into a computational format [see Sections 2.2, and 3.2.1].

iii. ML data analysis: different ML algorithms were tested to analyse the data. In relation to the data used, the accuracy and sensitivity response of the ML algorithms were also measured [see Chapters 8, 9, Section 2.2]. More details regarding the associated stages can be seen in Section 3.3.2.

vi. Data verification: the data will be trained and tested in two sub-modes for the demonstrator (i.e., descriptor, predictor) and trained in one sub-mode (i.e., descriptor) for the prototype. Additionally, the sensitivity response will be tested on the descriptor mode [see Chapters 8, 9], and the predictor for the demonstrator only [see Chapter 8]. The applicability of the model entails contrasting the influence of the different rules (e.g., sub-modes, accuracy) with actual data. The data was split into four different sample ratios (i.e., 50/50, 80/20, 85/15, 75/25). With a view to exploring which set would provide the best possible extraction of new knowledge to contribute to this study, the examination was carried out based on the four ratios above mentioned and seven risk factors explored in this study (Lanera et al., 2019; Saha et al., 2021)[see Chapters 2, and 8].

Risk Factors

In the context of this research study, a risk factor is described as any contribution that increases the probability of a road-crash fatality. In addition to that, it is statistically related to crash frequency and severity (Elvik et al., 2009). Subsequently, these risk factors can affect events before, during, and after a road-crash period (WHO, 2018). In this study, risk factors were investigated with the view to identifying patterns that may support the anticipation of pedestrian road-crash fatalities (Elvik et al., 2009; WHO, 2015). It is also important to note that, even though road-crash casualties result from a series of events affected by many risk factors within a traffic system, in this study, only a limited number of road attributes were considered. Therefore, to assess pertinent information on pedestrian general attributes resulting in casualties, the two following components are suggested:

i. To examine the definition of some specific elements in a road environment (e.g., pedestrian gender, pedestrian location, road condition, weather) [see Sections 2.1 and 2.2].

ii. To assess and validate more accurate modelling techniques (e.g., unsupervised-learning) that can discover new knowledge through data [see Section 2.2].

3.3.2. Pedestrian Road-Crash Attributes

According to many in the field, pedestrian road data can be used to improve pedestrian safety [see Section 2.2]. The road data availability from the STATS19 database, is developed and advanced by the Department for Transport (DfT) in the UK. Additionally, it is made available to provide an overall measure of road safety. In the context of preventing pedestrian road-crash casualty, three main categories were identified: pedestrian general attributes, road environment and road infrastructure (Haddon, 1970) [see Section 2.2]. These were also identified on the STATS19 form [1]. In this study, some of the attributes that might be taken into consideration are

[1] Appendix A.1

as follows:

i. Pedestrian general attributes (e.g., Gender, Age group).

ii. Road environment (e.g., Speed Limits, Weather Condition, Lighting Condition).

iii. Road infrastructure (e.g., Road Type, Pedestrian Physical facilities).

The selected attributes to be used in the computational datamining tool, can be represented by their respective headers. For example, on pedestrian gender: male, and female. These variables are designated as a subset of the attributes, i.e., metadata, which are used to summarise basic information and enable easier tracking when working with specific attributes (Addakiri et al., 2019). This applies to search for specific attributes, in the event of a pedestrian casualty in a particular location (e.g., roundabout). Other attributes that can be represented by the computational tool are described as follows: *road type: in a carriageway crossing pedestrian facility; in a carriageway crossing within zigzag lines at crossing approach; on footway or verge; in a carriageway crossing elsewhere; on refugee, central island or central reservation* . A similar principle applies to trace any other type of attributes (e.g., weather, road condition) (DfT, 2017).

Machine-Learning: Data Analysis Stages

Various research studies acknowledged pedestrian safety as a dynamic and complex topic (Naumann et al., 2019) [see Sections 2.1, 2.2]. The literature review findings on pedestrian safety analysis indicated that, compared with traditional statistical methods, the ML methods provide more objective results [see Chapter 2]. For this reason, the latter was adopted.

Figure 3.4.: ML data analysis diagram

Figure 3.4 represents the six stages carried out on the ML data analysis. A descriptive overview is presented below:

Stage 1. *Data Gathering*: To collect pedestrian road-crash data from public sources (e.g., STATS19) (DfT, 2017; DfT, 2018; DfT, 2019). To unify the collected data into a single format (i.e., attributes), using a specific computational datamining tool (KNIME, 2021) [see Chapters 3, 6].

Stage 2. *Data Filtering*: To limit the pedestrian road-crash casualty data to a region within country (e.g., GB in the UK) or a city (e.g., Enfield in London). To limit data to answer any exploratory research question (e.g., pedestrian casualties in a road environment) (DfT, 2017).

Stage 3. *Algorithm Selection*: to select a data-learning algorithm (i.e., ML algorithm) based on the pedestrian and road attributes under analysis.

Stage 4. *Algorithm Assessment*: to assess the data-learning algorithm accuracy, by comparing different data algorithms regarding their accuracy in relation to different profiles of data (e.g., 1-year).

Stage 5. *New Rules Validation*: to verify validate the data-learning algorithm applicability considering different sub-modes (i.e., descriptor, predictor) into the actual road dataset.

Stage 6. *Implementation:* experimental results were carried out: to test the data-learning algorithm sensitivity response in relation to different profiles of data (e.g., local, regional). To analyse the results that were be presented in a mathematical and/or a graphical form. To contrast the findings from the datamining analysis (e.g., pedestrian behaviour, road infrastructure) with the literature review and validate these with expert opinion.

The above procedure was carried out using a computational datamining tool which is usually designated as relational. In most cases, this type of structure allows one to identify, group, and access data through organised tables, divided into meaningful sub-tables (e.g., table identification) (Hong et al., 2020). A further explanation will be described in Chapter 6.

3.3.3. Data quality: Filtering, Cleansing, Verification - The Konstanz Information Miner (KNIME)

In raw databases as STATS19, the quality of the data can be affected by hardware failures, repeated entries, and wrong values (Embrey, 2002). To improve data quality the main steps undertaken were filtering, cleansing, and verification. Data filtering addressed selecting relevant variables (i.e., road attributes) for the analysis, data cleansing consisted of discarding undesired content from databases by removing the errors, support filling in the missing fields, and rectify inconsistencies (Embrey, 2002) [see Section 3.3.2]. Additionally, data verification regarded exploring performance metrics to extract results were split into train, and test modes (Buscema, 2013) [see Section 2.5].

Data Filtering

Table 3.1 presents some of the selected variables for the examination procedure.

Table 3.1.: Variables selection from STATS19

Attributes	**Included**	**Excluded**
Accident Index	X	
Number of Fatalities	X	
Casualty Class	X	
Casualty Severity	X	
Car Passenger		X
Bus or Coach Passenger		X
Pedestrian Location	X	
Casualty Gender	X	
Age Group of Casualty	X	
Speed Limits	X	
Road Type ([00.] No Junction (20 m); [01.] Roundabout; [02.] Mini-Roundabout; [03.] Staggered Junction; [05.] Slip Road; [06.] Crossroads; [07.] Other Junction; [08.] Private Drive; [09.] Other Junction)	X	
Pedestrian Facilities ([0.] No Crossing Facility (50 m); [1.] Zebra; [4.] Pelican or Similar; [5.] Ped Phase at traffic signal junction ; [7.] Footbridge or Subway; [8.] Central Refuge)	X	

Table 3.1 exemplifies the selection both the included and excluded variables from STATS19 to KNIME for the examination (KNIME, 2021) [see Section 6.2.1]. Since this research only focuses on pedestrians in a road environment, the variables regarding the pedestrian's location (e.g., Pedestrian_Location) and the severity of the accidents (Casualty_Class) are relevant. For instance, to examine the elderly age group, data only involving people with 60 years old or above can be selected. Regarding the physical infrastructure, the road type and the pedestrian physical facilities were included. However, even though the information is extracted from the same source, the numbering system between the road type and the pedestrian physical facilities differ (e.g., from 01 to 1). Out of the scope of this study are vari-

ables that relate to other type of road users (e.g., Car_ passenger, Bus or Coach_ passenger).

Data Cleansing

Figure 3.5 illustrates an example of the pedestrian profile data cleansing procedure.

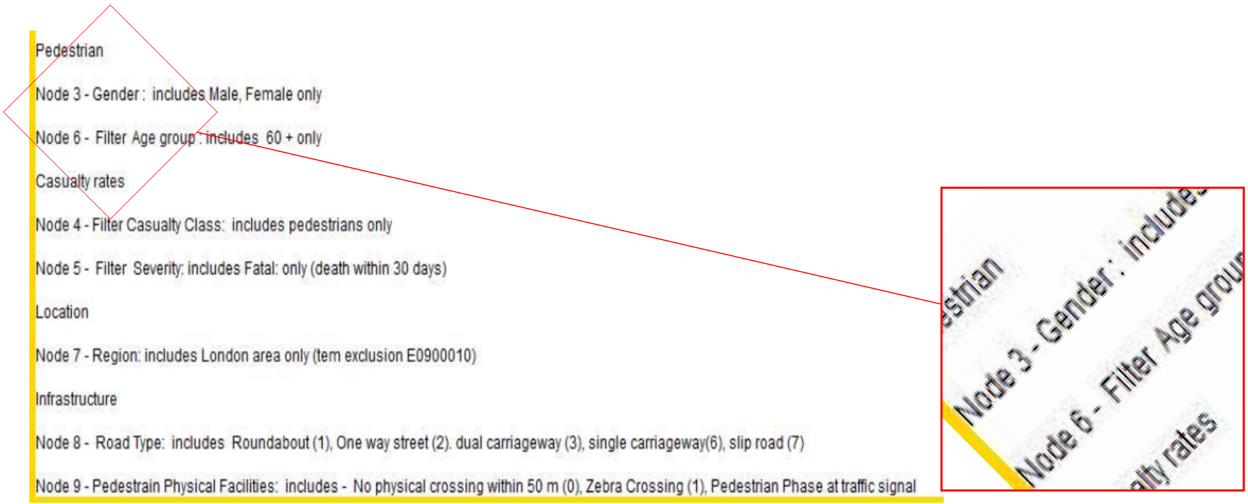


Figure 3.5.: Variables content cleansing: gender, age group

Figure 3.5 illustrates the undertaken procedure to discard undesired content from the selected variables. Some of the examples of the Nodes included in the analysis are as follows:

Node 3: Males and females (i.e., 1 and 2). All the other variable records (e.g., gender: -1) are discarded.

Node 6: On the age groups, unknown data (i.e., age group: unknown)

Figure 3.6, illustrates an additional example of the data cleansing procedure on the road infrastructure which includes discarding undesired content from the selected variables (e.g., road type, pedestrian physical facilities) from STATS19.

Figure 3.6.: Variables content cleansing: road type, pedestrian physical facilities

Figure 3.6 highlights the content records cleansing on Node 8 (i.e., road type) and Node nine (i.e., pedestrian physical facilities). On Node eight, all the other records regarding road type records (e.g., -01, 04) are discarded. About Node nine, other records on pedestrian physical facilities (e.g., -1, 3, 6) are discarded. In addition to that, a similar procedure was repeated on the remaining nodes. This means that, with respect to pedestrian general attributes, an analysis on Nodes three (i.e., gender), four (i.e., Casualty Class), five (i.e., severity) and six (i.e., age group) were also cleared and predefined. Surprisingly, it was noticed that between the road type and pedestrian physical facilities, the numbering system differs (e.g., 01 vs. 1). On a wider scale, it is important to point out that this type of inconsistency may lead to undesired errors when forming conclusions.

Data Verification

Figure 3.7 shows the data verification procedure.

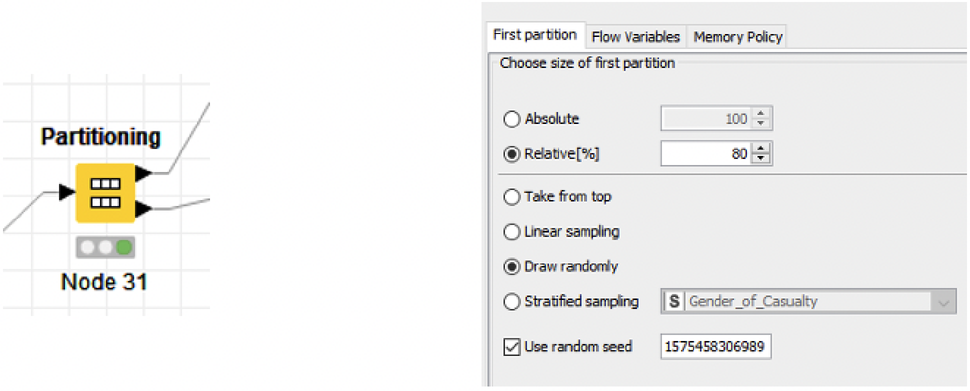


Figure 3.7.: Data verification: training, testing (KNIME, 2021)

Figure 3.7 illustrates data verification procedure in a set of performance metrics initially split into two datasets: train and test [see Section 2.5]. The training dataset, i.e., descriptor, to build the model (e.g., 80%), the test dataset, i.e., predictor, to validate the model (20%) correctness (Lanera et al., 2021; Saha et al., 2020). To get reproducible results upon re-execution, the type of data partition was performed using a random seed where different models were tested and compared using a common set of data split (Fiorentini and Losa, 2020; KNIME, 2021). Moreover, the response sensitivity evaluation of the selected algorithm was tested using different partition scenarios (e.g., train: 70%, and test: 30%). These will be introduced in Section 6.3.3 and explored in Chapter 9.

3.4. Evaluation

The evaluation process consists of 3 main functions: model assessment, newly discovered knowledge, and conclusions. In general terms, this means that from comparing the results of the evaluation process with the literature review, as the new knowledge discovered might likely be applied to similar road environments (Hoy et al., 2015) [see Section 2.5; Chapter 9].

3.4.1. Model Assessment

The accuracy and objectivity of the ML models are respectively validated and verified using scoring metrics. In addition to that, the suitability of the data under study and the capability of the computational datamining tool are assessed. As part of the model assessment, the verification and validation of distinct ML algorithms (e.g., supervised, and unsupervised) was examined in more detail using different sets of data [see Chapter 7]. First, the accuracy of the selected model was performed (Abdar et al., 2021 , Maryam et al., 2023) [see Section 2.1.1]. After, the model validated through the adoption of a ranking method (e.g., co-relational statistics) (Sarraf and McGuire, 2020) [see Section 2.5]. About the model verification, upon defining their modes and their sets of rules, both their accuracy and sensitivity response were tested [see Chapters 8, 9].

3.4.2. System Implementation: Experimental Results

The analysed attributes regarded a number of pedestrian general attributes (e.g., gender, level of impairment) and also infrastructure elements (e.g., speed limits, physical facilities) [see Chapters 4 and 5]. Regarding predictive modelling, a number of supervised and unsupervised ML algorithms were explored in Chapter 7. The prototype investigated the applicability of an ML algorithm to discover unknown trends, such as links resulting from integrating distinct pedestrian road-crash elements. The prototype will be demonstrated in Chapter 8 and its sensitivity testing response, will be presented in Chapter 9. To extract knowledge from a computational datamining tool, multiple experimental simulations were carried out. In addition to that, the prototyping highlighted comparing results between knowledge extracted from the computational datamining tool and information from additional sources [see Section 2.5, Chapter 8]. For its validation, it is highly important to note that the findings might be strictly applicable to the road-crash data under this study (e.g., year, region) [see Chapter 10].

3.4.3. Knowledge Discovery

An illustration of the knowledge discovery process is presented in Figure 3.8.

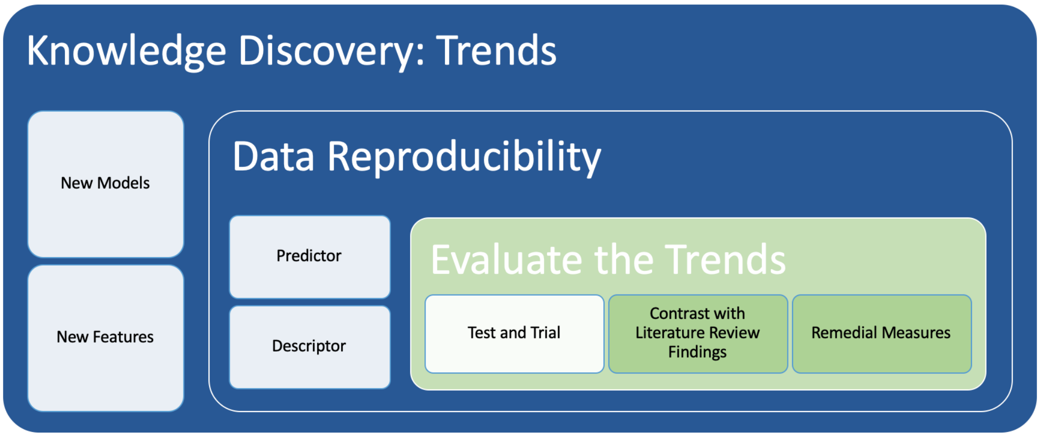


Figure 3.8.: Knowledge discovery

The data under study in Figure 3.8 was analysed as follows:

i. Road data from existing public databases.

ii. ML road data analysis.

iii. Comparative analysis between ML road data, previous research work, and expert opinion.

Moreover, the data discovery was developed under the Knowledge Discovery Database (KDD) principles. Generally, the KDD concept is associated with discovering useful knowledge through data (Shapiro, 2020). The ML data analysis is applied to KDD to explore and assess existing pedestrian safety gaps regarding unknown data links and patterns to date. Moreover, it is believed that different ML methods (e.g., unsupervised-learning) can be proposed to improve pedestrian safety. One example involves the incorporation of new datasets, enhancing the models' validity by improving accuracy (Bastien et al., 2020). In terms of data verification, different sub-modes and rules of the ML technique may assist in identifying information requiring further attention that is unknown to date (Elton, 2018). Subsequently, a discussion on the results ensued, contrasting the experimental findings with those identified in the literature review and the expert opinion [see Chapters 2, 8, 9, and 10].

3.5. Observations

Recent advancements in the field of pedestrian safety highlight the challenges posed by disregarding known causes of pedestrian road-crash casualty causation in an urban road environment (Stigson et al., 2008; Elvik et al., 2009; Cordellieri et al., 2019; Nhac-Vu et al., 2014; WHO, 2018; Ang et al., 2020; DfT, 2020; Jagnoorand Peden, 2020; Giummarra et al., 2021). Moreover, to date there has been limited consensus on standardising modelling approaches (e.g., descriptive, predictive) towards pedestrian safety challenges. Since an urgent need to address this issue remains, to overcome traditional forecasting techniques, a discrete classification applying ML algorithms was examined (Yannis et al., 2013; Feliciani et al., 2020; Giummarra et al., 2021) [see Section 2.2]. The chapter presented the adopted ML methodology to investigate attributes causing pedestrian fatalities in an urban road environment, spatially (i.e., 2017 road data from GB and London) and temporally (i.e., 2009-2019). The undertaken approach, proved to reveal undiscovered patterns with a lower dependency on historical data, and on human intervention. Thus, a more independent method known as 'self-learning' (Elton, 2018; Xiupeng, 2019; Bastien et al., 2020). However, due to the small size of the dataset, it was not possible to precisely determine the relationship between the current road data (i.e., 2017) and the subsequent data (i.e., ahead in time) [see Section 8.2]. The accuracy was calculated using two metrics: margin of error (MOE) and standard score (z-score) (Tanur, 2011; Sakai et al., 2015; Uddin et al., 2023). Another key issue may be that the results lack relevance due to restricting the analysis to the data under study. A distinctive advantage of this method is its potential for expansion, especially when relating data from different regions [see Section 7.2.4]. It is therefore likely that such an approach will bring an opportunity to more proactively support road engineers in discerning the influence of pedestrians on a given road infrastructure while steering further research on pedestrian safety [see Section 2.6]. The next chapter examines GB pedestrian road-crash casualty data by adopting engineering logic to investigate a selection of pedestrian general attributes.

4. Pedestrian Road-Crash Casualties

4.1. Introduction

This chapter describes how pedestrian attributes can impact on pedestrian safety, in an urban road environment. The main purpose is to prove the applicability of predictive modelling to identify patterns from examining pedestrian attributes. The chapter is organised in three sections. Section 4.1 introduces the overall structure of the chapter. Section 4.2 examines a number of existing predictive road-crash models and their applicability according to STATS19 through engineering logic. It presents a summary of the discovered knowledge from the investigated methods, and it also describes their relevance (e.g., advantages, limitations). Section 4.3 discusses the findings.

4.1.1. Overall Approach

Worldwide and in Europe the pedestrian road-crash trends differ per country (WHO, 2018; EU STATS, 2019; DST, 2020) [see Section 1.2]. Traditionally, statistical analysis are suggested to be performed in cross-sectional studies (Cahill, 2010; Wisdom and Creswell, 2013) [see Section 2.2]. However, over a longer period using purely data-driven information it might not be feasible to identify patterns (e.g., outdated model) (Urie et al., 2016; Useche et al., 2019). This seems to be highly critical. In addition to that, an underlying degree of weakness remains which is clearly specifying reasons contributing to pedestrian road-crash casualty causation. A few of the shortcomings could be due to some of the information being omitted, overwritten, overlooked, or even misperceived (Chen and Fan, 2019). The mentioned shortcomings could be a likely reason for a much-debated question on pedestrian general attributes affecting pedestrian safety. It has been previously reported that, to draw more conclusive results, patterns, and trends on pedestrian safety more research is needed (Wisdom and Creswell, 2013; DfT, 2018; Useche et al., 2019). An example is Cahill (2010) suggesting the incorporation of mixed research methods (i.e., qualitative and quantitative) as an instrument to advance a more systematic integration to examine pedestrian safety (Wisdom and Creswell, 2013). Therefore, in this study a mixed research method was adopted (Gonzalez-González et al., 2021; Raifman et al., 2021) [see Sections 2.2, 2.6]. With respect to the examined pedestrian general attributes in Chapter 2, this section presents a few previously used pedestrian safety models by gender and age group. Additionally, their limitations in terms capabilities will be described.

4.2. Pedestrian Predictive Modelling

Frequent changes in a complex traffic road environment including population groups, time periods, and also geographic regions seem to lead to road-crashes [see Section 2.3]. Predictive models have demonstrated to make an efficient interpretation with an effect to discern quick changes. Moreover, compared with traditional models which are reliant on statistical models (e.g., forecasting), predictive models which rely on ML models have demonstrated their ability to project road-crashes more accurately and also more objectively (up to 99%) [see Section 2.2]. In this section, these will be explored based on a set of criteria from STATS19 as the gender (i.e., female, male) and the age groups (i.e., children: 0-15 years old; young adolescents: 15-24 years old; adults: 25-59 years old; elderly: 60+ years old) (DfT, 2019; OECD, 2020) [see Sections 2.2.1 and 2.3.1].

4.2.1. Gender: Females, Males

To predict the gender of the pedestrians in the event of a road-crash casualty, the ordered logit model,

$$P(Y_{\mathrm{i}} = 1|X_{\mathrm{i}}) = F(X_{\mathrm{i}}\beta) \tag{4.1}$$

Source: (Trueck and Rachev, 2009)

Wherein the accident severity codes according to STATS19 are given on a scale as follows: i equal to one is a fatal accident, equal to two is an accident with serious injury, and equal to three is an accident with a slight injury (DfT, 2017). The model selection emerges for its a binary classification: female, male (Trueck and Rachev, 2009). P is the gender probability that Y*(0=female; 1=male)* [see Section 2.2]. X represents the gender predictor variable. F(.) the continuous random variable logistic distribution of the model and β describes the relationship between the gender prediction variable and the prediction response (Allen and Singh, 2011). One key affecting issue is the dependability from data set recognition and the reliance on police crash reports considering evidence from different data sources. Scalability is another crucial limitation. For example, discrepancies noted between the data sample (i.e., 2756) where fatalities reached 1.8% of accuracy, and the weighted sample (i.e., 326, 452) where fatalities rates reached 0.2% of accuracy) (Kaplan and Prato, 2012).

4.2.2. Age Group: Children

To predict the children (0-15) pedestrian fatalities in the event of a road-crash casualty, the negative binomial model,

$$ln(\lambda_{\mathrm{i}}) = B_{\mathrm{X_j}} + \epsilon, \tag{4.2}$$

Source: (Graham et al., 2005)

In which the wards of England are spatial units with an average area of 14 km^2. *i,j* are the distances from ward i to ward j. λ_{i} is the expected number of children pedestrian fatalities per ward, and $_{\mathrm{X_j}}$ a vector of regressors (Graham et al., 2005). β is the estimated coefficient with the expected number of children pedestrian fatalities per period from a unit change in an independent variable X_j (Graham et al., 2005; STATS19). One crucial issue is that the model does not clarify variation in child or adult fatalities. To achieve such, a highly detailed data analysis of very small and restricted samples is required. Therefore, the results from this model are usually recommended as complimentary (White et al., 2000).

4.2.3. Age Group: Young Adolescents

To predict the young adolescents (15-24) pedestrian fatalities in the event of a road-crash casualty the logit model,

$$Logit(P(Y=1|x_{\mathrm{i}}....,x_{\mathrm{k}})) = B_0 + B_{\mathrm{i}}.X_{\mathrm{i}} + ... + B_{\mathrm{n}}.X_{\mathrm{n}} \tag{4.3}$$

Source: (Trueck and Rachev, 2009, pp. 22)

Where P is the probability of an accident to occur or not: [0; 1]. B_0, B_1,...,B_n which are the model constant variables (i.e., the occurrence of the accident) the parameter estimates for the independent variables, X represents the independent variable (e.g., age of the pedestrians) and i is the set of the severity of the accidents (i = 1, 2,........., n), and $\left(Logit_{\mathrm{P}}\right)$ is the the natural logarithm ranges from negative infinity to positive infinity. One key issue is the validation in the use of default models (Trueck and Rachev, 2009; Wedagama and Dissanayake, 2020). Using a single factor seems to be insufficient to predict the occurrence and severity of road crashes. A key challenge might be to identify the optimal number and type factors to be modelled (Mohanty and Samal, 2019).

4.2.4. Age Group: Adults

To predict adult (25-59) pedestrian fatalities in the event of a road-crash casualty, the multinominal model,

$$P_{\text{in}} = \frac{exp(\beta_{\text{i}} X_{\text{n}})}{\sum_j exp(\beta_{\text{j}} X_{\text{n}})} \tag{4.4}$$

Source: (Chen and Fan, 2019)

Where P_{in} is the probability of pedestrian β_i is the probability of crash n with injury level $_i$, X_n is a vector of explanatory variables that determine the severity. The accident severity codes according to STATS19 are given on a scale as follows: i equal to one is a fatal accident, equal to two is an accident with serious injury, and equal to three is an accident with a slight injury (DfT, 2017). β_i, represents a vector of estimable coefficients for the injury outcome. One issue affecting the model, is its high dependency on the variables. Over time this might likely affect the quality of the model (e.g data fluctuation, data loss). Thus, the validity of the model predictions in terms of accuracy and transferability is likely to be affected (Chen and Fan, 2019).

4.2.5. Age Group: Elderly

To predict the elderly (60+) pedestrian fatalities in the event of a road-crash casualty, the logistic regression model,

$$P(Y = 1|X) = \frac{exp(b_0 + \sum_{i=1}^{n} b_{\text{i}} x_{\text{i}})}{1 + exp(b_0 + \sum_{i=1}^{n} b_{\text{i}} x_{\text{i}})} \tag{4.5}$$

Source: (Sanz-Casado et al., 2019)

Where P(Y=1| X) is the probability that Y takes the value one. The dependent variables are classified as: '0=severe injury accident group (accidents that list seriously injured individuals but no deaths)' and '1=fatal accident group (accidents with at least one fatality)'. X is the set of n co-variables (e.g., gender, age group, impairment level) from X_i to X_n that are part of the model. b_0 is the model constant or independent term. b_1 is the coefficient of the co-variates. This predictive model was applied to a final sample consisting of 225 accidents, considering 27 variables. One main disadvantage was the data sample restriction. Therefore, exploring alternative methods (e.g., Poisson regression) to examine other human factors and additional tuning parameters was suggested (Sanz-Casado et al., 2019). However, it is important to note the risk of overfitting models both in terms of data reliability and accuracy (Niebuhr et al., 2019).

4.2.6. Knowledge Discovery: Pedestrian General Attributes

Table 4.1 illustrates a summary of the examined predictive models by gender and age group.

Table 4.1.: Predictive pedestrian models summary

Pedestrian Attributes	**Predictive Models**	**References**
Gender		
Female, Male	*Ordered logit: see* 4.2.1	(Kaplanand Prato, 2012)
Age Groups		
Children	*Negative binomial: see* 4.2.2	(Graham et al., 2005)
Young Adolescents	*Logistic regression: see* 4.2.3	(Wedagamaand Dissanayake, 2010)
Adults	*Multinominal logit: see* 4.2.4	(Chenand Fan, 2019)
Elderly	*Binary logistic: see* 4.2.5	(Sanz-Casado et al., 2019)

Table 4.1 summarises the examined predictive models by gender (i.e., binary logit) and age group (negative binomial; logistic regression; multinominal logit; predictive

simulator). About their key noted limitations, the binary logit failed to recognise data (e.g., police road data) (Niebuhr et al., 2019). This suggests potential measurement errors leading to data discrepancies. The negative binomial showed that only with detailed or restricted data samples it would be possible to distinguish the age groups which is limiting for schools with a larger number of students with a wider range of age groups. This proved lack of validity regarding scalability. Therefore, the results from negative binomial are usually recommended as complimentary (White et al., 2000). The logistic regression proved a validation constraint to identify the optimum variables (e.g., number of factors, type) when using default models (Sanz-Casado et al., 2019). Similarly, the multinominal logit identified a high dependency on the selected variables (Chen and Fan, 2019). More research to confirm if selecting variables through more complex rules is suggested (Niebuhr et al., 2019). As a complementary method to increase reliability, the use of the binary logistic is suggested (Sanz-Casado et al., 2019). Turning into ML method applied in this research, differently from previous examined predictive models, it consists of using a singular predictive modelling approach to extract knowledge from pedestrian attributes in an urban road environment. Respectively by gender and age group as follows:

- Case studies on pedestrian risk perception, pedestrian distraction and pedestrian impairment.

- Descriptive modelling applied on pedestrian general attributes (DfT, 2018, KNIME, 2021).

- Validation of the findings with expert opinion.

Subsequently, the description of patterns regarding undiscovered attributes influencing pedestrian safety will be presented. This exercise is demonstrated in the form of a prototype in Chapters 8 and 9.

4.3. Observations

The investigation of previously applied pedestrian safety predictive models by gender, age group, showed various limitations (e.g., lack of data recognition, scalability). The adopted ML examines the same number of pedestrian attributes by means of a singular analysis. This will be demonstrated in Section 8.3. However, it's important to note that human error can be a fundamentally influential factor on the examination of pedestrian road-crash casualty [see Sections 2.3, and 2.3.1]. Regarding pedestrian road-crash casualty by gender and age group, the literature review indicated the female, elderly, and children as the most exposed pedestrian to road casualties [see Sections 4.2.1, 4.2.2, and 4.2.5]. Moreover, about measuring different levels of pedestrian impairment, the investigated case studies reveal limitations in terms of risk awareness [see Sections 4.2.1, and 4.2.2]. As regards qualitative research on the different pedestrian cognitive disabilities, a limited body of literature was found (Uzan and Wagstaff, 2017; Earl et al., 2018; Hezaveh and Cherry, 2018) [see Section 2.3]. In this study, the adopted mixed-method (i.e., qualitative: case studies and expert opinion, and quantitative: pedestrian safety models) points to a number of limitations [see Sections 2.6, and 4.2.6]. For instance, the investigated case studies could have been more comprehensive if they had incorporated predictive datamining analysis to explore the level of impairment of the pedestrians (Robbins, 2023; Tabone and De Winter, 2023). This approach is out of the scope of this study. About the quantitative research, the restriction on the number of participants and their location per case study, seems to be a critical disadvantage for scalability purposes [see Sections 2.2, 2.6, and 2.3.1]. Moreover, examining predictive models confirmed a number of limitations: the high dependency on historical data, selection of default models and variables, and also the models expiry date [see Sections 2.2, 2.6, and 4.2.6]. One example was the predictive simulator being limited to prognose shortcuts wherein young and elderly pedestrians could likely be positioned [see Chapter 2]. Therefore, to expand the new knowledge acquired from this study, further research on predictive mixed methods to investigate the pedestrian general attributes influence on pedestrian safety is suggested (Oxley et al., 2005; Nemire et al., 2016; Earl et al., 2018; Tabone and De Winter, 2023). Overall, the findings revealed the validity of scrutinising information regarding pedestrian general attributes. However, this chapter is limited to examine the pedestrian general attributes that affect pedestrian safety (Hatfield and Murphy, 2007; Twisk et al., 2015). A number previously applied predictive models only accounting road elements that affect pedestrian safety, will be described in greater detail in the next chapter.

5. Road Infrastructure

5.1. Introduction

This chapter shows the influence of the road infrastructure on pedestrian safety. The purpose of this chapter is to prove the applicability of previously applied predictive modelling to identify patterns from examining road attributes, in an urban road environment. The chapter is organised in three main sections. Section 5.1 introduces the overall approach of the chapter. Section 5.2 investigates, through engineering logic, a number of existing road predictive models, while demonstrating their relevance in an urban road environment. Section 5.3 discusses the findings.

5.1.1. Overall Approach

The World Road Association (WRA) (2015) points to road infrastructure as a significantly influential factor on fatal pedestrian road-crash causation. Previous research studies have suggested the improvement of the road infrastructure performance (e.g., road lighting systems, road classification) as it helped reduce road-crashes notably (Elvik et al., 2009) which subsequently, reduced the costs of road-crashes (Adminaite et al., 2015; VfR, 2015). Multiplicative factors, also known as Crash Modification

Factors (CMF) have been used to reflect the expected changes in pedestrian safety performance (Elvik et al., 2009; Chen and Persaud, 2013). In Europe, the most recent Accident Prediction Models (APMs) and crash modification factors (CMF) are available in a public database. In 2016, the total number summed 273 APM's and 889 CMF's respectively (PRACT, 2016). According to Papadimitriou et al., 2015, understanding relations between the road infrastructure and measurable CMF improved pedestrian safety (CMF Clearinghouse, 2013; iRAP, 2020; Austroads, 2018). In the same vein, the absence of pedestrian facilities has been considered as a major road-crash influential factor, especially at illegal street crossings (Bungum et al., 2005). Turning to GB roads, in 2017, the collected data show pedestrians reaching 470 fatalities in total where they account the road infrastructure as follows:

- Road infrastructure (i.e., road type): intersections 27.87 %; roundabouts 3.62 % (DfT, 2019) [see Section 2.4.5].

However, the current analysis on GB road data from 2017 does not show results regarding the influence of other factors such as the weather, speed limits and pedestrian physical facilities (DfT, 2017; DfT, 2018) [see Section 2.4]. Moreover, Section 2.6 and Chapter 4 point to the expiration validity dates of the predictive models as a crucial constraint. This chapter explores the applicability of existing predictive models using STATS19 road data to discover information from existing road elements influencing on pedestrian safety (Cahill, 2010; DfT, 2020).

5.2. Road Infrastructure Predictive Modelling

The road environment and its layout prove to considerably influence the safety of the pedestrians (Cameron and Nilsson 2004; Yannis and Karlaftis, 2010; Yannis et al., 2013, Kangas et al., 2015) [see Section 2.4.5]. To demonstrate the relevance of existing predictive models in GB, the set criteria is based on STATS19 and include weather; speed limits; lighting; T or staggered junction; roundabout; private drive entrance; pedestrian facilities: no physical facilities within 50 m, zebra, footway, central refuge (DfT, 2017; OECD, 2020).

5.2.1. Weather

To predict road-crash casualty based on weather (e.g., fine, raining, snowing, fog), the integer autoregressive model,

$$y_{\mathrm{t}} = a_{\mathrm{t}} \circ y_{\text{t-1}} + \epsilon t, t = 1 ... T \tag{5.1}$$

Source: (Yannis and Karlaftis, 2010)

Where ϵt is a Poisson time sequence with E $(\epsilon t) = \lambda > 0$ and independent of t-1. Each component either 'survives' with probability a or not (1 - a). The function introduced by Van Harn and Steutel (1977), is given by a $\circ y \sum_{i=1}^{y} u_{\mathrm{i}}$

Where u is a sequence of binary random variables. A limitation of this model is the number of variables taken into account for the analysis: vehicle accidents, vehicle fatalities, pedestrian accidents and pedestrian fatalities, were the type of pedestrian facility is not represented. Additionally, it requires a large set of historical data: 21 years of daily count data (Yannis and Karlaftis, 2010).

5.2.2. Speed Limits

To predict the road-crash casualty based on speed limits, a mathematical model,

$$\left(\frac{Killedpedestrians_{\text{after}}}{Killedpedesdrians_{\text{before}}}\right) = \left(\frac{Meanspeed_{\text{after}}}{Meanspeed_{\text{before}}}\right)^4 +$$

$$+\left(\frac{Meanspeed_{\text{after}}}{Meanspeed_{\text{before}}}\right)^8 \times \left(Numberoffatalities_{\text{total}} - Numberoffatalcrashes_{\text{total}}\right) \tag{5.2}$$

Source: (Nilsson, 2004; Nilsson, 2010)

Nilsson (2004) has developed a polynomial mathematical model using several variables that contrast the relationship between changes in mean traffic speed, the number of crashes and killed pedestrians. An example is that, if assumed that 265 pedestrians fatal crashes occur in a year, and a total of 300 people are killed as a result from those crashes, at a ratio of 0.9 (i.e., mean speed after to mean speed before), the predicted pedestrian fatal crashes (PPFC),

$$PPFC = 0.9^4 \times 265 = 0.656 \times 265 \approx 174 \tag{5.3}$$

A decrease on the number of fatal crashes by more than 34% is projected. With respect to the predicted pedestrian fatalities (PPF),

$$PPF = (0.9^4 \times 265) + (0.9^8 \times (300 - 265)) \approx 174 + 15 = 289 \tag{5.4}$$

A 37% decrease in the number of pedestrian fatalities resulting from road crashes is projected. The results on this cumulative model with fitting exponents that apply to killed pedestrians (i.e., PPF), is higher than the exponents that apply to crashes (i.e., PPFC) (Nilsson, 2010). One key limitation is the use of the model when there is a need to represent more than three dimensions. This applies if, for example, a raised intersection is required to be represented as a countermeasure (King et al., 2003). To overcome this shortcoming, manual work was performed.

5.2.3. Lighting

To predict pedestrian fatalities based on road lighting, the Mean Absolute Percentage Error (MAPE),

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{F_t} \right| \tag{5.5}$$

Source: (Marino et al., 2017)

where A_t is the actual traffic value, F_t is the forecast traffic value and N is the number of observations. To determine the traffic a week ahead and evaluate the data reliability, statistical methods (e.g., regressions and neural networks based on algorithms) have been adopted. MAPE is used as a calculation to predict the accuracy of these various statistical models, through comparison and to improve the public lighting system in an urban road environment. Regarding the applied training technique, the MAPE data was split into two sets of data: train 75% (about 4800 records) and test 25% (about 1600 records). The study was tested for 12 months in a local street named Strinella, in L'Aquila, Italy (Marino et al., 2017). The model accuracy improvement might likely require to both incorporate more sophisticated statistical models and to train the data over a longer period.

5.2.4. Intersections

To predict pedestrian fatalities at an intersection the negative binomial regression model,

$$N_{\mathrm{Ped}} = exp(\beta_1 ADT + \beta_2 PedVol + \beta_3 X3 + ... + \beta_{\mathrm{n}} X_{\mathrm{n}}) \tag{5.6}$$

Source: (Torbic et al., 2010)

where $\beta_0 to \beta_{\mathrm{n}}$ are the coefficients to be estimated, $\mathrm{N}_{\mathrm{Ped}}$ is the expected number of pedestrian crashes, ADT is the average daily traffic (i.e., vehicular volume), PedVol is the average daily pedestrian volume, and X_3, X_{n} are other site characteristics, as proportion of the total volume that is left turn, number of lanes, speed limit, presence or absence of a crosswalk, and presence or absence of a median. Therefore, through the results of this model, information on the island refuge could also be obtained. However, on account of low pedestrian road-crash frequency, a large sample of sites with many years of data is needed. Lyon and Persaud (2002), applied the model in 122 intersections in the three-leg intersections and compiled 11 years of data at the locations. Torbic et al. (2010) collected data from 2,000 locations.

5.2.5. Roundabouts

To predict pedestrian fatalities at a roundabout a regression model,

$$C = (4.56Peds + 2.00VPH - 3.00Dist).10^{-4} \tag{5.7}$$

Source: (Stone et al., 2002)

Where C is the predicted annual pedestrian-vehicle crashes, Peds is the pedestrian flow on one leg of the intersection in the peak hour, VPH is the conflicting vehicle flow in the peak hour, Dist is the maximum street crossing distance. On the latter, the first three factors showed the highest correlation to pedestrian road-crash. This model has been tested at the Hillsborough-Horne intersection, in Raleigh, USA. Nevertheless, the scarcity of pedestrian road-crash data at roundabouts has been considered critical. Therefore, alternating approaches to assess pedestrian safety at roundabouts as among others, carrying out case study analysis, statistical analysis, and simulation analysis has been previously suggested (Stone et al., 2002; Cahill et al., 2010).

5.2.6. Pedestrian Facilities: Zebra

To predict pedestrian road-crash at zebra crossing the ordered binary logit model,

$$U_{\mathrm{i}} = ai + \beta i1X1 + \beta i2X2 + \beta i3X3 + \beta i4X4 + ... + \beta inXn \tag{5.8}$$

Source: (Al Bargi and Daniel, 2020)

Wherein U_{i} is the utility of choosing alternative facilities, i is the alternative (i.e., use/not use); n is the number of independent variables, a is a constant, b are the coefficients. The probability of choosing alternative i,

$$P(i) = \frac{1}{1 + exp(Ui)} \tag{5.9}$$

Source:(Al Bargi and Daniel, 2020)

The utility of alternative i must be transformed into a probability in order to predict whether a particular alternative will be chosen or not. The findings of the experiment indicated that the option of using the pedestrian crossings was influenced by more than one variable (Al Bargi and Daniel, 2020). However, interestingly none of the identified ones were related to the design of the road infrastructure.

5.2.7. Pedestrian facilities: Present vs. Absent

According to iRAP (2015), the Number of Predicted Accidents per year (PNAY),

$$ENAY \times CMF = PNAY \tag{5.10}$$

Wherein ENAY is the number of accidents experienced per year, CMF are the crash modification factors (i.e., pedestrian crossing a road in an urban road environment with no facility), PNAY is the predicted number of accidents per year. Assuming that,

i. ENAY is equal to 200 and CMF is equal to 6.7 (i.e., no facility), then PNAY is equal to 1340.

ii. ENAY is equal to 200, and CMF is equal to 0.4 (i.e., grade separated facility), then PNAY is equal to 80.

The result of the first calculation indicates that as result of not implementing the countermeasure (i.e., no pedestrian facility), 1340 accidents a year are predicted. By contrast, the second calculation indicates that because of implementing a countermeasure (i.e., grade separated facility), 80 accidents a year are predicted (iRAP, 2015; Turner et al., 2017).

5.2.8. Knowledge Discovery: Road Infrastructure

Table 5.1 illustrates a summary of the examined road infrastructure predictive models.

Table 5.1.: Road infrastructure predictive models summary

Infrastructure	**Predictive Models Description**	**References**
Road Environment		
Weather	*Integer autoregressive model: see* 5.2.1	*(Yannis and Karlaftis, 2010)*
Speed Limits	*Mathematical chain ratio see:* 5.2.2	*(Nilson 2004; Nilson 2010)*
Lighting	*MAPE see:* 5.2.3	*(Marino et al., 2017)*
Road Infrastructure		
Intersection	*Negative Binomial Regression see:* 5.2.4	*(Zegeer et al., 2020)*
Roundabouts	*Regression see:* 5.2.5	*(Stone et al., 2002)*
Zebra	*Ordered binary logit see:* 5.2.6	*(Al Bargi and Daniel, 2020)*
Pedestrian Facilities: present vs. absent	*CMF see:* 5.2.7	*(Turner et al., 2009)*

Table 5.1 summary shows that predictive models for the pelican, footbridge and private property were not found. Regarding the key issues on the integer autoregressive model and ordered binary logit, these are restricted to the number of the variables. To improve the accuracy of these models, an enhanced approach including additional or more relevant variables was suggested (Yannis and Karlaftis, 2010; Al Bargi and Daniel, 2020). About the mathematical chain ratio demonstrated limitations in presenting images in three dimensions. A crucial issue is human intervention, as it can lead to increasing errors (e.g., misspelling) (Nilsson, 2010). Both the MAPE and Regression models are limited to the location of experimentation. To increase reliability in different regions, the adoption of these methods as complementary might be appropriate. (Stone et al., 2002; Cahill et al., 2010; Marino et al., 2017). With respect to the negative binomial regression, none of the identified variables were related to the design of the road infrastructure. In addition, large sample of sites with many years of data are required (Lyon and Persaud, 2002; Torbic et al., 2010). Turning into ML method applied in this research, similar to the approach presented in Section 4.2.6, a singular predictive modelling approach to extract combined knowledge from road environment infrastructure is as follows:

- Case studies on weather whether, speed limits, lighting, intersections and pedestrian physical facilities.

- Descriptive modelling applied on the road environment and road infrastructure characteristics (DfT, 2018, KNIME, 2021).

- Validation of findings with expert opinion.

Subsequently, the description of patterns regarding undiscovered road attributes influencing pedestrian safety will be presented. This exercise is demonstrated as a prototype in Chapters 8 and 9.

5.3. Observations

The results in this chapter recognises road elements as significant influential factors to improve pedestrian safety (iRAP, 2021; Stigson et al., 2009) [see Section 2.4]. Therefore, looking to a few environmental and infrastructural features to identify the most relevant road elements affecting pedestrian safety is crucial. Similar to Chapter 4, the findings from the explored environmental (i.e., weather, speed limits, lighting), infrastructural factors (i.e., road intersections, pedestrian physical facilities) and also the predictive models, identified various limitations (e.g., restricted to the number of variables, variables unrelated with the design of the road) [see Chapter 5]. With respect to road environment, in terms of weather, the implementation of physical facilities lacks accounting the weather influence (e.g., rainfall) [see Section 2.4.2]. Regarding speed limits, variable speed limits and traffic calming treatments (e.g., gates, raised intersections, raised crosswalks) reduces pedestrian road-crashes [see Section 2.4.3]. On lighting, higher illuminance increases the probability of detecting obstacles, reassures the pedestrians cognitive judgements. Turning now to the road infrastructure, the results indicated the presence of pedestrian facilities as a primary factor to reduce pedestrian road-crashes [see Section 2.4.4]. However, unfortunately on some of the pedestrian physical facilities such as the pelican, footbridge and private property, predictive models were not found. The advantage of adopting ML in this study is for its ability to explore the unforeseen links between human, environmental and infrastructural attributes on a singular analysis. Moreover, for the pedestrian physical facilities where no predictive models have been deployed to date, these can also be included in the analysis. This will be demonstrated in Section 8.3. However, similar to previously deployed pedestrian predictive models confirmed to have some limitations: high dependency on historical data, and selection of relevant variables (Papadimitriou et al., 2019) [see Sections 2.3.1, 2.6, 4.2.6 and 5.1.1]. Regarding data reliability and model validity, the incorporation of predictive methods can likely face challenges (e.g., accuracy, and expiry date, specified location) (Cahill, 2010; Zangenehpour et al., 2015). Moreover, a shortcoming of the investigated methods is the level of uncertainty on the applicability, and scalability to other regions, owing to lack of data quality (e.g., data discrepancies). Furthermore, it has been thought that the validity of the investigated models could have offered a more comprehensive exploratory approach. This is because more reliable results might be obtained through combining qualitative (e.g., case studies) and quantitative research (e.g., surveys on road surface condition, pedestrian facilities). Subsequently, more convincing conclusions could have been drawn [see Section 2.4.5]. Therefore, in this study a comparison between findings of the predictive modelling and qualitative studies was carried out (Bungum et al., 2005; Oxley et al., 2005; Earl et al. 2018; Nemire et al., 2016; Jagnoor, 2020) [see Section 2.6]. Nevertheless, this chapter is limited to road elements that affect pedestrian safety (Hatfield and Murphy, 2007; Twisk et al., 2015) [see Section 2.3]. The next chapter will present a predictive modelling approach, which combines data regarding pedestrian attributes and road elements to examine pedestrian road-crash casualties.

6. The Analysis Process: Data Exploration

6.1. Introduction

This chapter introduces a systematic approach to investigating the applicability of predictive models on actual road data. The proposal delineates the significance of ML data analysis for discovering new knowledge using a datamining tool called KNIME. This tool enables the integration of various components including ML algorithms, for data examination. Additionally, it enables performing verification and validation techniques. The objective of this data analysis is to demonstrate the existing advantages of the adopted method which integrates data from STATS19 about pedestrian attributes and road elements, respectively expanded in Chapters 4 and 5 (DfT, 2017). The data analysis was deployed to explore previously undiscovered patterns or links affecting pedestrian safety [see Section 2.4]. This chapter is organised in five main sections. Section 6.1 shows the overall structure of the chapter. Section 6.2 details the data process analysis which entails the following two items:

i. The definition of road features from STATS19.

ii. The association,filtering and verification of road data using KNIME.

Section 6.3 presents the data model overview and clarifies its selection and assessment procedures. Section 6.4 presents a typical examination of actual road data from STATS19 using KNIME. Section 6.5 discusses the findings.

6.1.1. Overall Approach

Recent evidence indicates that lack of harmonised data sources, may be a critical aspect interfering with the ability of predictive models to conclude more realistic results (Deb et al., 2017, Assi et al., 2020). The importance and originality of this study lies in its exploratory and interpretative nature, as the adopted ML method consists of exploring STATS19 data regarding pedestrian attributes and road elements. Additionally, prior to selecting the most relevant ML data algorithm, the analysis follows a pre-defined data exploration procedure, including assessment and validation. Its applicability will be demonstrated in Chapter 8.

6.2. Data Exploration: Road-Crash Casualty in GB

As regards the data examination, the description of the road attributes provides a basis for defining features from STATS19 and transferring these to KNIME. Moreover, the data association allows for linking the transferred features to KNIME and converting them into a single dataset format. In this study, a spatial analysis using ML and a temporal analysis using traditional methods were carried out [see Chapters 8, 10 and Section 3.2.2]. The proposed process to explore GB pedestrian road-crash casualty data, is displayed in Figure 6.1.

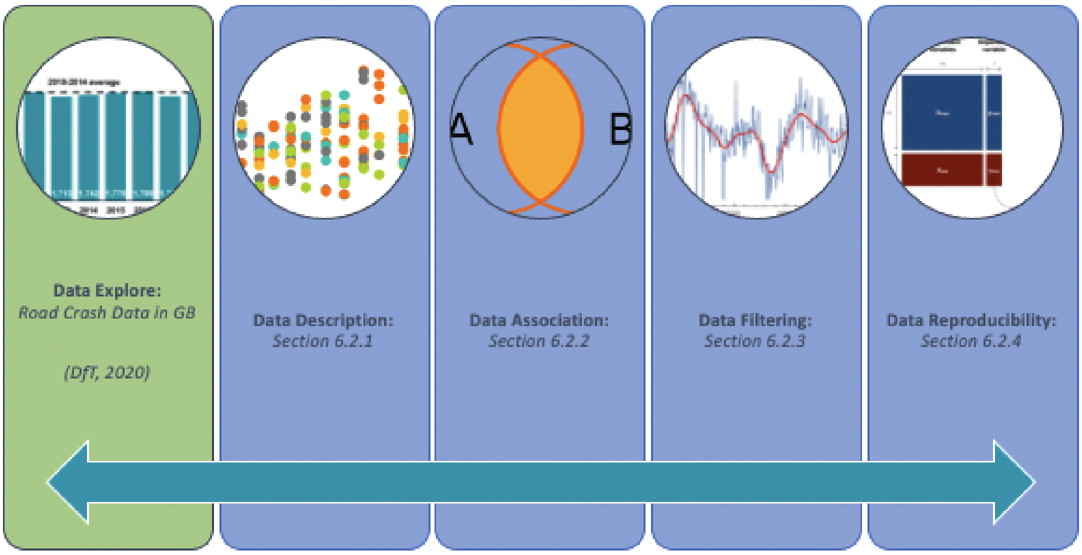


Figure 6.1.: Road-crash data exploration process

Figure 6.1 illustrates the process of exploring data divided into 4 main stages: attributes description, data association, data filtering, data verification (KNIME, 2021). The data has been curated (i.e., filtered and cleansed) using the computational database KNIME [see Section 3.3.2]. In this process, seven attributes (i.e., variables considered useful for the analysis) were gathered. Moreover, to improve data quality, the data filtering process discards both undesired variables (e.g., vehicle_reference) and their respective undesired content (e.g., gender: unknown,

-1). Also, there are a number of discarded attributes from the STATS19 form (e.g., Pedestrian Crossing – Physical facilities: -1, 3, 6) (Appendix 11.9). Furthermore, a data verification technique was applied. It is important to highlight that the developed model has been broadened to select the most relevant ML algorithm for the data under study. A number of existing ML algorithms are further described in Chapter 7.

6.2.1. Road Data Attributes: STATS19, KNIME

KNIME is written in Java and based on Eclipse and is compatible with being executed on other software languages such as Java, Python, R, and Ruby. Additionally, it supports an incremental compiler to provide additional functionalities, as follows:

Native connectors: SQLite, MS-Access, SQL Server, MySQL, Oracle, PostgreSQL, and Vertica and H2.

Data Transformation: Filter, Converter, Splitter, Combiner, Joiner, and ROC.

Data Configuration: Add, and Remove.

Visualisation: Report Designer, .svg, and .png.

Import/Export: .doc, .ppt, .xls, .pdf, .svg, and .epf.

Figure 6.2 illustrates the procedure for the transfer of road data from STATS19 to KNIME.

Accident_Index
Location_Easting_OSGR
Location_Northing_OSGR
Longitude
Latitude
Accident_Severity
Number_of_Casualties
Date
Day_of_Week
Time
Local_Authority_(District)
1st_Road_Class
1st_Road_Number
Road_Type
Speed_limit
Junction_Detail
Junction_Control
Pedestrian_Crossing-Human_Control
Road_Surface_Conditions
Special_Conditions_at_Site
Carriageway_Hazards
Urban_or_Rural_Area
Did_Police_Officer_Attend_Scene_of_Accident
LSOA_of_Accident_Location
Vehicle_Reference
Casualty_Reference
Casualty_Class
Sex_of_Casualty
Age_of_Casualty
Age_Band_of_Casualty
Casualty_Severity
Pedestrian_Location
Pedestrian_Movement
Pedestrian_Road_Maintenance_Worker
Casualty_Type
Casualty_Home_Area_Type
Casualty_IMD_Decile

(a) Variables of the road attributes

Row ID	S Accident_Index	I Vehicle...	I Casualt...	I Casualt...	S *Gende...	I Age_of...
Row0	2017010001708	1	1	2	2	18
Row1	2017010001708	2	2	1	1	19

(b) Road data attributes content

Figure 6.2.: STATS19 data from road attributes (KNIME, 2021)

Figure 6.2 shows the analysed road attributes from STATS19 as follows (DfT, 2017) [see Chapter 2]:

i. Pedestrian general attributes (i.e., gender, age group).

ii. Road environment (i.e., weather, speed limits, lighting).

iii. Road infrastructure (i.e., road type, pedestrian physical facilities).

Regarding the data features, these were first extracted from STATS19 using accidents and casualties. The selected road data was defined by dependent and independent variables. The dependent variables concerned the total pedestrian road-crash casualties, whereas the independent variables regarded attributes such as pedestrian profile, road environment, and road infrastructure. Figure 6.2a presents the dependent variables that resulted from combining STATS19 information from both the accidents file, and the casualties file. These variables are the metadata summarising road-crash information to ease the track when working with specific attributes (e.g., age, gender) [see Section 3.3.2]. The content of the variables presented in Figure 6.2b can be either numeric (e.g., integer, double) or nominal (e.g., string). This regards information from the metadata age (e.g., 17 years old), and gender (e.g., 2 - female). Furthermore, based on a set of possibilities, the type of the road attributes was pre-specified (e.g., integer) and named per category (e.g., year: 2017). In this study, a combination of numeric and nominal attributes is presented as names or labels. The data association process used in KNIME will be described in the next section. In terms of pedestrian physical facilities, as a central road element in this study, the following types were further described: zebra, pelican, central refuge, footbridge or subway, private driveway [see Chapters 2, and 5].

6.2.2. Data Association - KNIME

Figure 6.3 illustrates the data association procedure with KNIME.

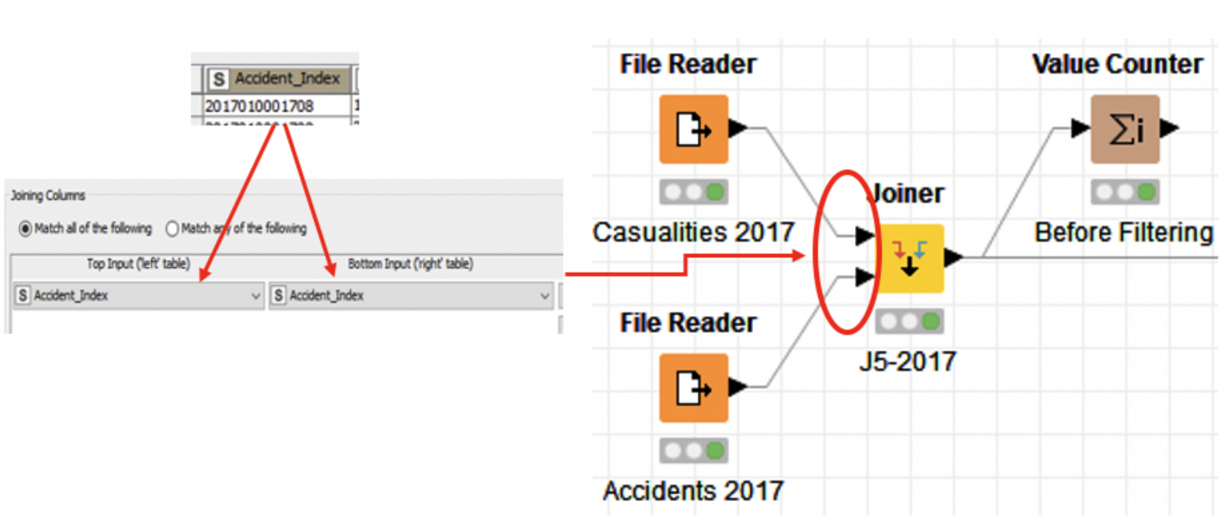


Figure 6.3.: Data association (KNIME, 2021)

Figure 6.3 shows the data association process to merge two separated files using distinct information (i.e., accidents and casualties) into a single format through a common AccidentID (i.e., Accident_Index). It exemplifies the spatial data examination accordingly, referring to the total number of accidents: *File Reader Accidents 2017* (i.e., fatal and non-fatal) and to the total number of accidents resulting in fatalities: *File Reader Casualties 2017* (i.e., fatal) (DfT, 2017; KNIME, 2021). The same approach applies to the data examined in two distinct regions: London and GB. This is later presented in Chapter 8.

6.3. Data Algorithm Description

Figure 10.1 displays the set-out stages to select the ML data algorithm.

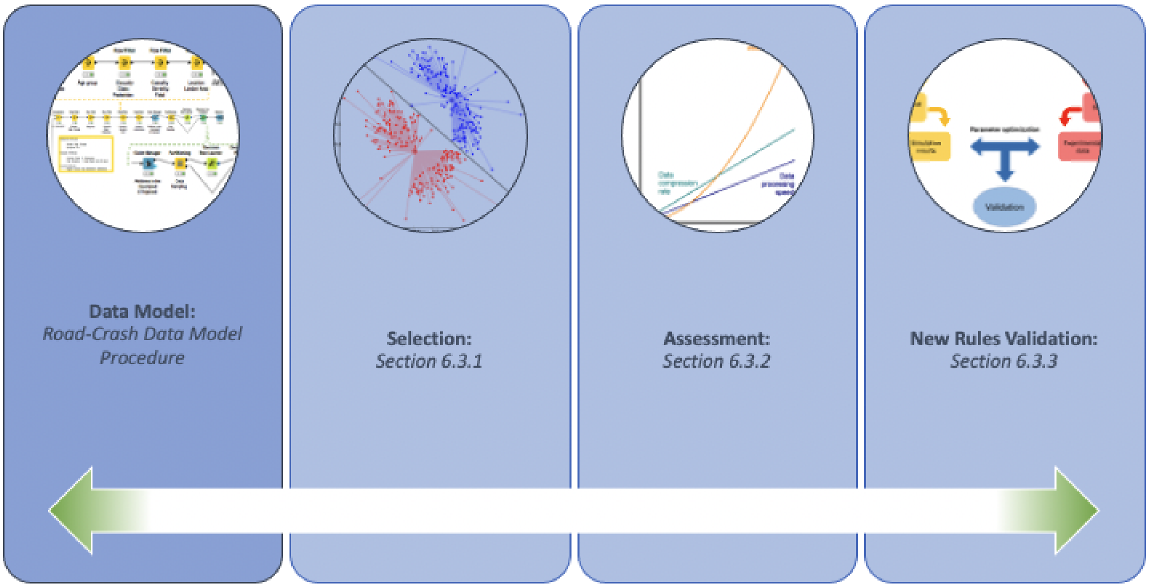


Figure 6.4.: ML data-learning algorithm procedure

Figure 10.1 illustrates the three steps carried out to deploy the ML data algorithm: selection, assessment, and new rules validation (KNIME, 2021) [see Section 3.3.2]. The first step regarded selecting the ML data algorithm (e.g., Decision-Tree, FCM). The second one reflected assessing the ML data algorithm based on a set of metrics of each of the selected models. The third step introduced new rules to assess the model accuracy when using actual road data from STATS19 (DfT, 2017).

6.3.1. Data Algorithm Selection

Various approaches have proven the suitability of the ML predictive algorithms to examine pedestrian road-crash data. However, a crucial challenge lays on identifying the most relevant and applicable ML algorithms for the data under examination. Figure 6.5, illustrates a general overview of previously adopted ML algorithms on pedestrian safety. Some of these algorithms will be later on further explored in Chapter 7.

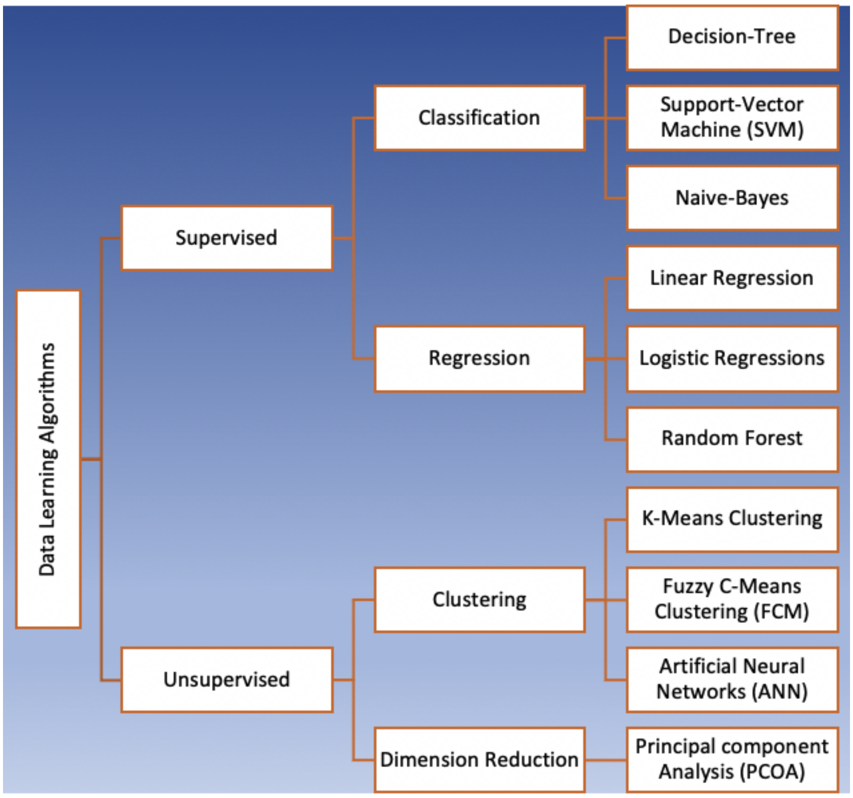


Figure 6.5.: ML predictive algorithms overview

Figure 6.5 presents a general overview of existing ML algorithms. Section 6.4 describes a traditional statistical data analysis using actual road data from STATS19 with KNIME. Chapter 8 presents the ML data analysis by applying the selected ML algorithm on actual road data from STATS19 with KNIME.

6.3.2. ML Data Algorithm Assessment: Quality Settings

To improve the quality of the findings when using data, on each of the selected ML algorithms two types of verification metrics were applied: Spearman's rho and Receiver Operating Characteristic (ROC). Whereas Pearsons' correlation is linear between two variables and when a change in one variable is associated with a proportional change in the other variable, the Spearman's rho correlation is a monotonic relationship between two variables (e.g., continuous, ordinal), where these tend to change together but not necessarily at a constant rate (De Winter et al., 2016). Spearman rho was used as a ranking method to assess correlations between the data attributes (e.g., road type vs. road surface condition) (Sarraf and McGuire, 2020). Generally, the accuracy metric is used to measure performance and quality of the algorithms (Silva et al., 2020; Abdar et al., 2021; Maryam et al., 2023). In this study, the ROC metric was used since it is more adequate to unevenly distributed data (Kwon et al., 2015; Silva et al., 2020). The ROC was adopted to assess the model in terms of its performance suitability and sensitivity response to the data under study [see Section 3.2].

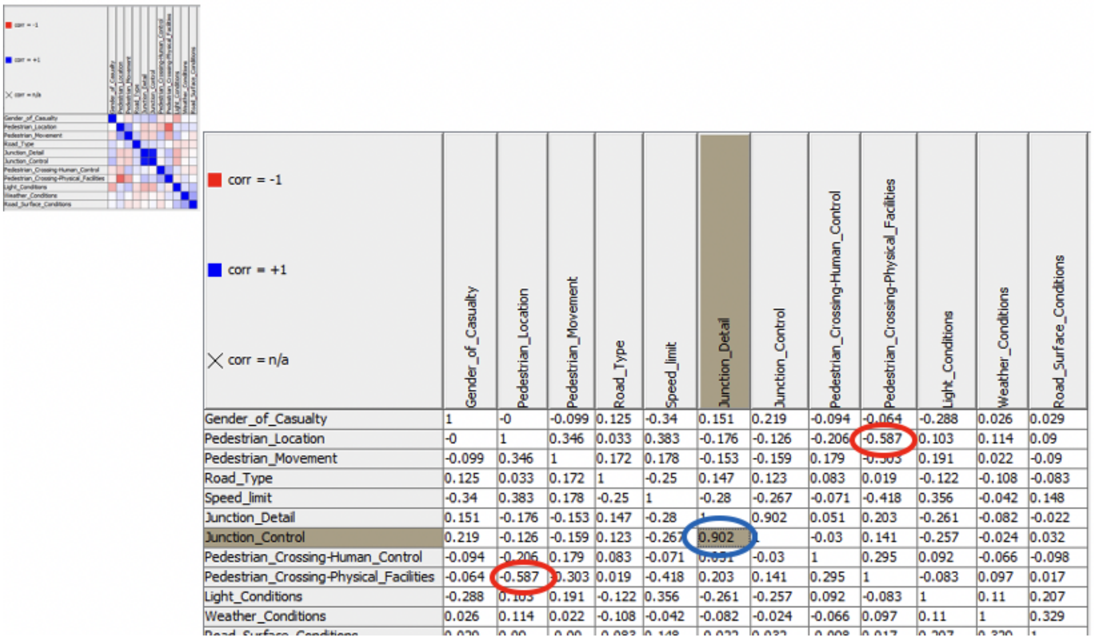
Spearman's rho

The Spearman's rho enables to rank the correlation between two different variables. The statistical calculation given by,

$$P(i) = 1 - \frac{6\sum_{d}{}^{2}{}_{\mathrm{i}}}{n(n^2 - 1} \tag{6.1}$$

Source: (Corder et al., 2014)

Where ρ is the Spearman's rank correlation coefficient, d_i is the difference between the two ranks of each observation and n is equal to the number of observations. The Spearman's rho ranking method was used to assess correlation between the data attributes (e.g., road type vs. road surface condition). Figure 6.6 shows the Spearman's rho technique ranking results when using KNIME.



	Gender_of_Casualty	Pedestrian_Location	Pedestrian_Movement	Road_Type	Speed_limit	Junction_Detail	Junction_Control	Pedestrian_Crossing-Human_Control	Pedestrian_Crossing-Physical_Facilities	Light_Conditions	Weather_Conditions	Road_Surface_Conditions
Gender_of_Casualty	1	-0	-0.099	0.125	-0.34	0.151	0.219	-0.094	-0.064	-0.288	0.026	0.029
Pedestrian_Location	-0	1	0.346	0.033	0.383	-0.176	-0.126	-0.206	-0.587	0.103	0.114	0.09
Pedestrian_Movement	-0.099	0.346	1	0.172	0.178	-0.153	-0.159	0.179	[illegible]	0.191	0.022	-0.09
Road_Type	0.125	0.033	0.172	1	-0.25	0.147	0.123	0.083	0.019	-0.122	-0.108	-0.083
Speed_limit	-0.34	0.383	0.178	-0.25	1	-0.28	-0.267	-0.071	-0.418	0.356	-0.042	0.148
Junction_Detail	0.151	-0.176	-0.153	0.147	-0.28	1	0.902	0.051	0.203	-0.261	-0.082	-0.022
Junction_Control	0.219	-0.126	-0.159	0.123	-0.267	0.902	1	-0.03	0.141	-0.257	-0.024	0.032
Pedestrian_Crossing-Human_Control	-0.094	-0.206	0.179	0.083	-0.071	[illegible]	-0.03	1	0.295	0.092	-0.066	-0.098
Pedestrian_Crossing-Physical_Facilities	-0.064	-0.587	-0.303	0.019	-0.418	0.203	0.141	0.295	1	-0.083	0.097	0.017
Light_Conditions	-0.288	[illegible]	0.191	-0.122	0.356	-0.261	-0.257	0.092	-0.083	1	0.11	0.207
Weather_Conditions	0.026	0.114	0.022	-0.108	-0.042	-0.082	-0.024	-0.066	0.097	0.11	1	0.329
Road_Surface_Conditions	0.029	0.09	-0.09	-0.083	0.148	-0.022	0.032	-0.098	0.017	0.207	0.329	1

Figure 6.6.: Spearman's rho (KNIME, 2021)

As can be seen in Figure 6.6, the Spearman's rho technique enables scoring every value of two nominal variables (e.g., pedestrian gender vs. pedestrian location). From analysing co-relational scores on road attributes between the junction control (e.g., automatic traffic signal, stop sign) and the junction detail (e.g., staggered junction, roundabout), the highest score peaked 0.902 (i.e., positive coefficient). By contrast, the lowest score reached -0,587 (i.e., negative coefficient). This resulted from contrasting pedestrian crossing facilities (e.g., zebra, pelican) and the pedestrian location (e.g., crossing on pedestrian facility in a carriageway, island refuge). The results indicate the automatic traffic signal at a staggered junction as highly compatible to examine pedestrian road-crashes. In contrast, the results suggest correlations between the pelican and the island refuge as least compatible to to examine pedestrian road-crashes. Figure 6.8 details information on the Spearmans' rho curvature, by are contrasting the pedestrian physical facilities with the pedestrian location.

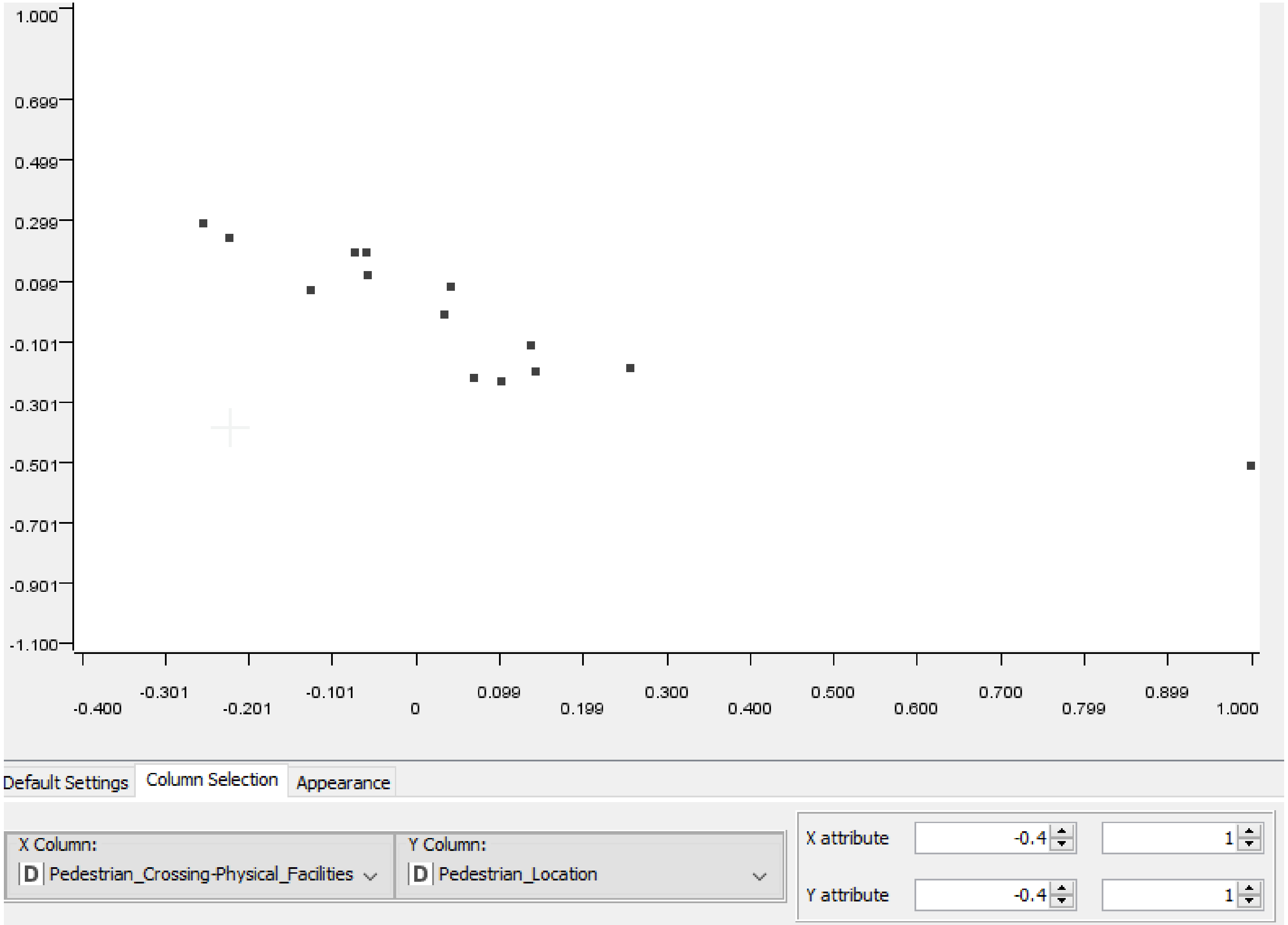
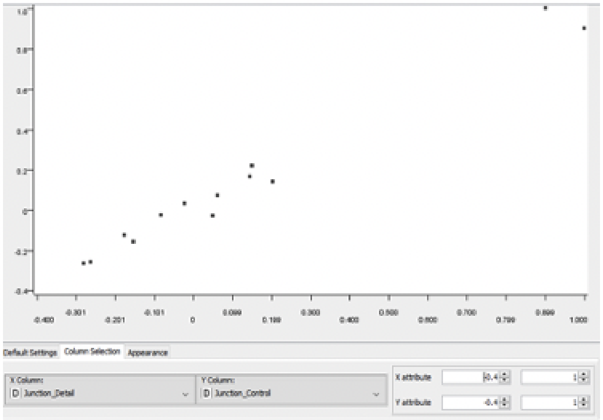
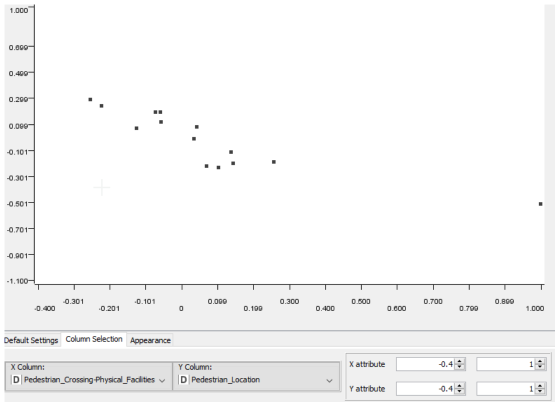


Figure 6.7.: Pedestrian physical facilities vs. pedestrian location

Moreover, after applying the Spearmans' rho, Figure 6.8 illustrates a contrast of two possible results (i.e., increase vs. decrease).



(a) Junction control vs. Junction detail



(b) Pedestrian facilities vs. Pedestrian location

Figure 6.8.: Spearman's rho curve: increase vs. decrease

The Spearmans' rho curve represented in Figure 6.8 increases in magnitude as the variables x (i.e., coefficients between -0.4 to 1.0) and y (i.e., coefficients between -0.4 to 1.0) become closer to being perfect monotone functions of each other. For example, a correlation type 1, occurs when the two variables are monotonically related. Even when their relation is non linear, i.e., nonparametric. When the two variables are perfectly monotonically related, the Spearmans' rho correlation coefficient becomes 1. Also, the closer the variables x and y are to each other, the more definitive data points appears [see Figure 6.8a]. On the other hand, in Figure 6.8b a correlation type 0 indicates no tendency for y to neither increase or decrease when x increases, here the data mostly appears roughly elliptically distributed.

Receiver Operating Characteristic(ROC)

Figure 6.9 illustrates the ROC curve.

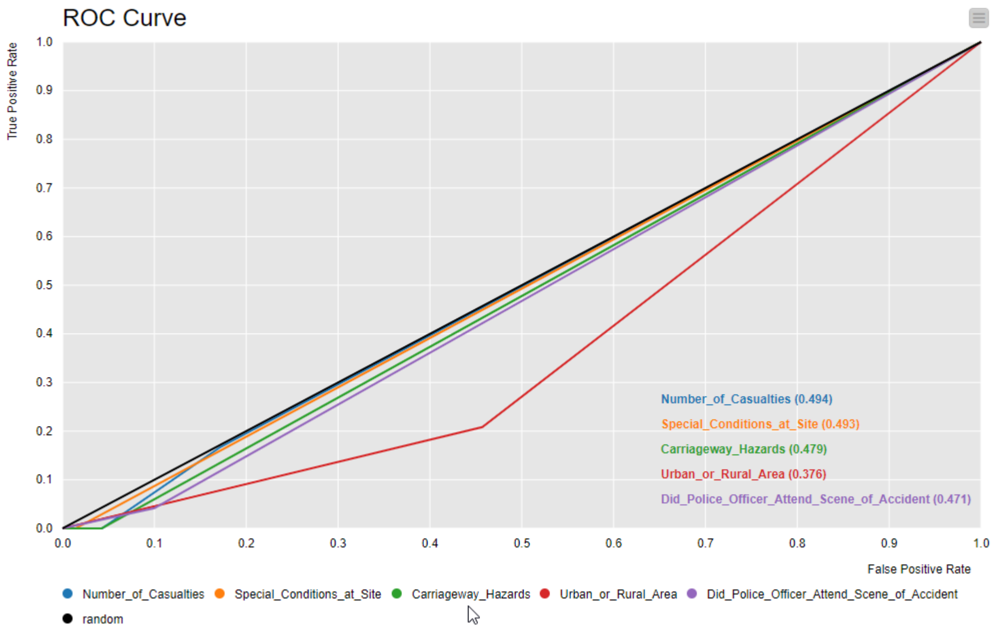


Figure 6.9.: ROC curve (KNIME, 2021)

An illustration of the ROC curve in Figure 6.9 is presented with a graphical plot. As a rule, it uses a random classifier which is the straight line from the origin (0.0, 0.0) to the top right corner (1.0, 1.0) presented diagram the black diagonal line (x, y) described as a random line. In general, a line that score between 0.5 and 1 (where 1 is the worse score and 0.5 indicates the algorithm is as good as the base model). On the other hand, for a line scoring below 0.5, it means the model performance best than the the base model (i.e., random classifier), which means being of no use. Moreover, the greater the area under the curve (i.e., x <0.5), the better the model sensitivity response. The ROC metric classifies real and probability attributes. Real attributes are classified as all possible values. The probability attributes classify the probabilities of real records. To create a ROC metric the attributes are sorted first by their class of probability, after for the positive class (i.e., rows sorted to front). Then, the sorted rows are checked to confirm if the real attributes belong to the positive class. Ideally, all positive results have a line going up to 1.0 on the true positive rate first. The results from the selected ML algorithm, score the highest on the urban or rural region (i.e., 0.698) and the least to identify the police presence where the accident occurred (i.e., 0.414). This means that with the selected ML algorithm, the ROC curve shows the least suitability to describe road-crash by region. On the other hand, it shows the best suitability to describe the police presence/absence in a road-crash. To diagnose the ML algorithm response sensitivity, the dataset from the descriptor and the predictor sub-modes can only be tested separately. Therefore, multiple ROC curves were applied to verify several scenarios for both sub-modes.

6.3.3. Data Correlation: New Rules Validation

The new rules are proposed to verify the ML algorithm applicability into the road dataset of this study considering the following parameters:

i. ML algorithm demonstrator: statistics 100%, descriptor 50%, predictor 50%.

ii. ML algorithm prototype: statistics 100%, descriptor 80%.

To assess the ML algorithm suitability when applied into the actual data from two distinct locations, the ML algorithm verification demonstrator will explore road data from GB, and London. Additionally, the ML algorithm prototype will verify the applicability on the overall GB road data. This will be later demonstrated in Chapter 8.

6.4. ML Data Analysis Advantages

6.4.1. Data Analysis Overview: GB

According to the STATS19 database, in 2017, in total 470 pedestrian road-crash fatality records were reported to the police (DfT, 2017). To assess road-crash casualties, KNIME explored actual data on pedestrian general attributes and road elements. As regards pedestrian general attributes, Figure 6.10 summarises the findings by gender.

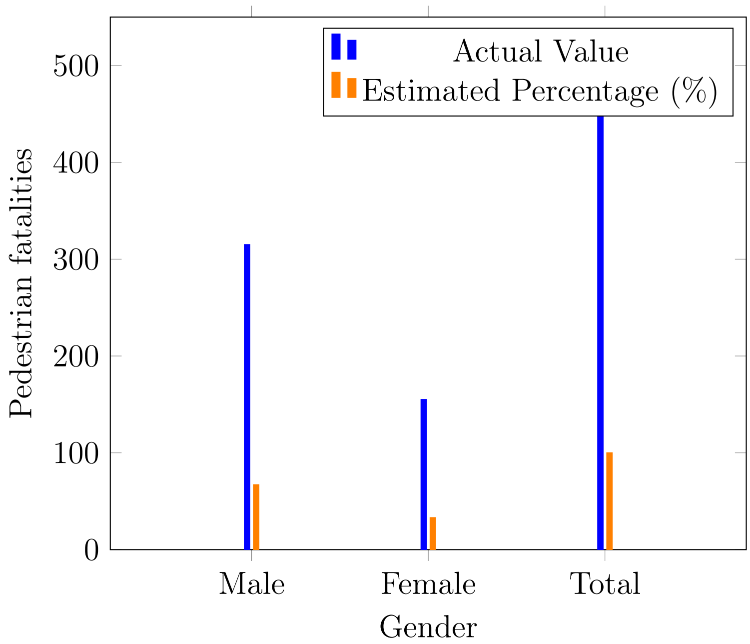


Figure 6.10.: Pedestrian road-crash by gender

The pedestrian attributes by gender in Figure 6.10, shows that males (67%) are nearly twice as much at risk than females (34%).

Figure 6.11, presents the findings by age group[1].

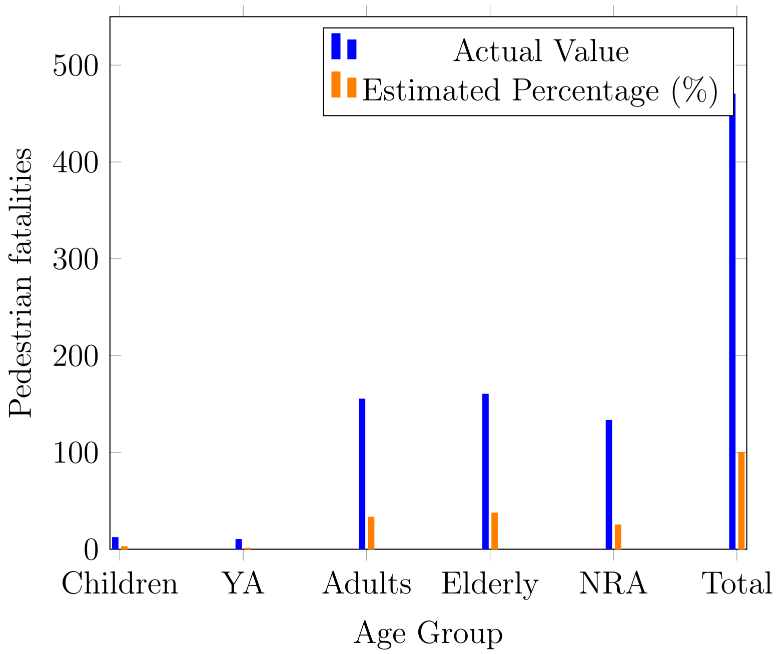


Figure 6.11.: Pedestrian road crashes by age group

As regards pedestrian attributes by age group, the elderly scored the highest (37.1%) followed by adults (34%). Interestingly, the NRA scored marking rates (25.3%). Additionally, a standalone dataset which is displayed below in Table 6.1 was also analysed.

[1] Children (0-14); Young adolescents (YA) (15-24); Adults (25-64), Elderly (65+); Non recorded ages (NRA)

Table 6.1.: Pedestrian road-crash: Standalone dataset

805	Dangerous action in carriageway (e.g., playing)
808	Careless, reckless or in a hurry
810	Disability or illness, mental or physical
803	Failed to judge vehicle's path or speed
802	Failed to look properly
806	Impaired by alcohol
807	Impaired by drugs (illicit or medicinal)
809	Pedestrian wearing dark clothing at night
804	Wrong use of pedestrian crossing facility

As shown in Table 6.1, pedestrian general attributes are not quantified. This may be due to data shortage. Still, the system proposes a hierarchic ranking. This was the case since *Dangerous action in carriageway* was ranked as first by KN-IME using Spearman's rho [see Section 6.3.2]. Previous research also confirms most road-crashes when pedestrians are randomly crossing the road. Moreover, *Careless, reckless or in a hurry* was ranked second. This can be interpreted as a significant point of unpredictability (e.g., distraction), which has been identified as another main reason to road-crashes (STATS19) [see Chapters 2 and 4].

Turning to the road environment, Figure 6.12 illustrates the findings by weather[2].

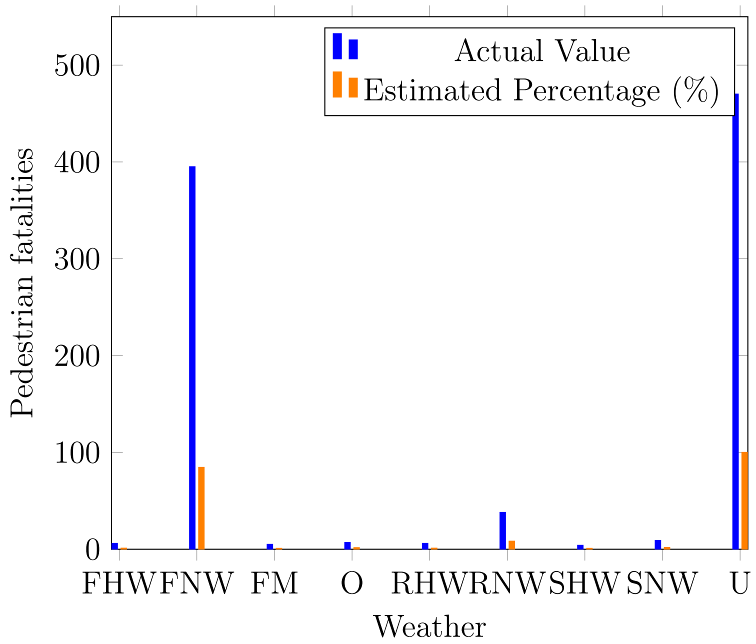


Figure 6.12.: Road-crashes by weather

Figure 6.12 shows the pedestrians at risk with remarkable rates when the weather is dry (i.e., 84.3 %), followed by when is raining (i.e., 8.26%).

[2]Fine-high winds (FHW); Fine-no winds (FNW); Fog or mist (FM); Other (O); Raining-high winds (RHW); Raining-no winds (RNW); Snowing-high winds (SHW); Snowing-no winds (SNW); Unknown(U)

Figure 6.13 shares the findings by speed limits[3].

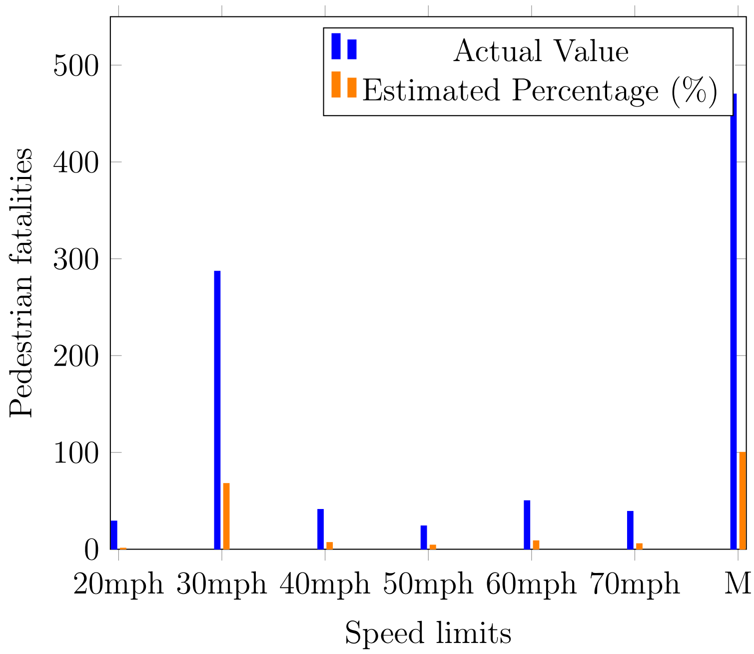


Figure 6.13.: Road-crashes by speed limits

Figure 6.13 presents that the pedestrians at risk up to 68% for permanent speed limits set at 30 mph (ca. 48.28km/h), respectively followed by 60 mph (ca. 96.56 km/h) with 8.61%, 40 mph (ca. 64.37 km/h) with 6.85% and 70 mph (ca. 112.65 km/h) with 5.6%.

[3] 1-20 mph, 21-30 mph, 31-40 mph, 41-50 mph, 51-60 mph, 61-70 mph, Motorway(M)

Figure 6.14 illustrates the findings by lighting condition[4].

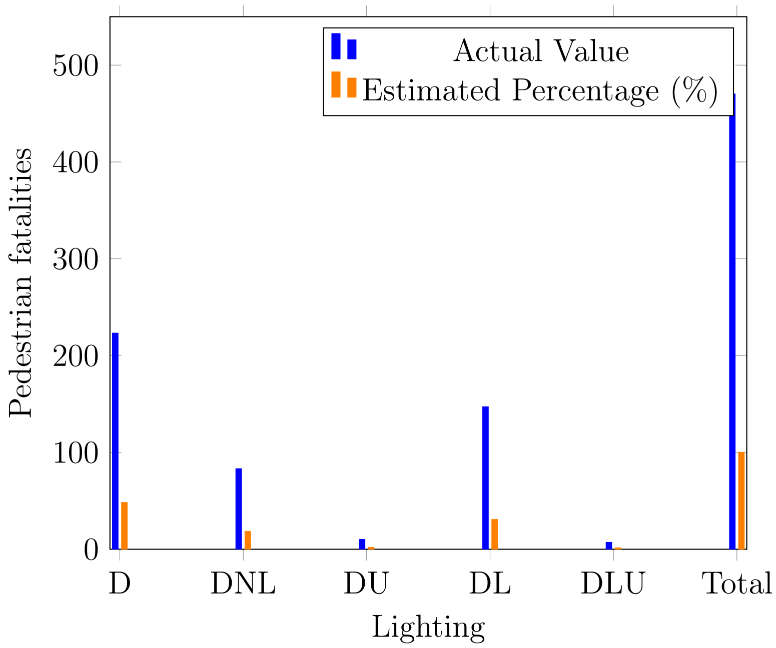


Figure 6.14.: Road-crashes by lighting condition

Regarding lighting, in Figure 6.14 data show the pedestrians at a higher risk (52.5%) during the day, followed by at night with light (30.6%), and at night without light (18.28%).

[4] Daylight (D); Dark no lighting (DNL); Dark unlit (DU); Dark lit (DL);Dark light unknown (DLU)

With regard to the road infrastructure influence on pedestrians, Figure 6.15 summarises the findings by road type[5].

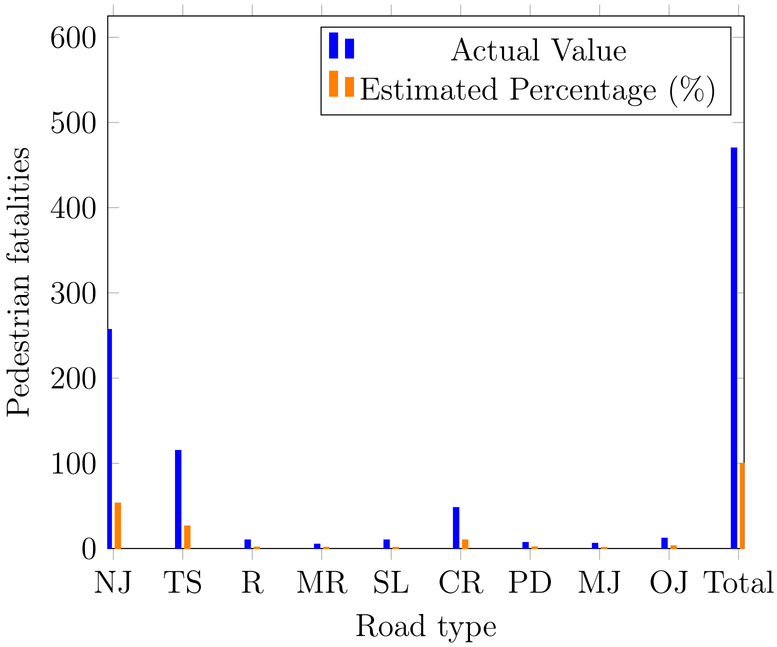


Figure 6.15.: Road-crash by road type

As can be seen in Figure 6.15, the pedestrians were at the highest risk when there were no physical facilities available, at no junction within 20 meters (53%), followed by when the pedestrians were located at a staggered junction (26.4%), and afterwards followed when the pedestrians were located at crossroads (9.97%).

[5]No Junction (NJ); T staggered (TS); Roundabouts (R); Mini-roundabouts(MR); Slip road (SL); Cross roads (CR); Private drive (PD) 7Multiple Junction (MJ); Other Junction (OJ)

Figure 6.16 presents the findings by the type of pedestrian physical facilities [6].

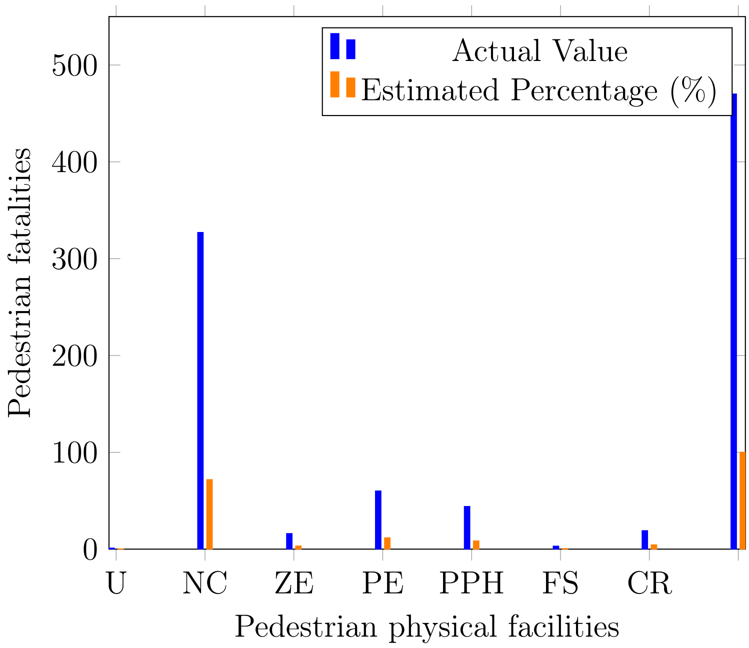


Figure 6.16.: Road-crash by pedestrian physical facilities

As per Figure 6.16, the absence of pedestrian physical facilities is significantly risky (68%), followed by the pelican (14%), and the Pedestrian Phase (8.44%).

6.4.2. Data verification Correlations

Regarding the spatial analysis carried out in 2017, GB, in total 470 pedestrian road-crash fatalities were analysed [see Section 3.2.2]. This number replicated respectively 376 (i.e., 80%) of trained data and 94 (i.e., 20%) of tested data. Interestingly, by contrast the standalone source in Table 6.1, traced in total 510 pedestrian road-crash fatality records. Subsequently, the data verification of this dataset (i.e., 80% vs. 20%) resulted in 408 of training in 102 of test data. One vital issue here is the difference between the two data sources by 40 records. This could happen due to, human intervention, and lacking on consistent database features or even on a standardisation on the data gathering procedure. Subsequently, similar to Section 3.3.3, when formulating conclusions, it could lead to question the validity of the undertaken analysis.

6.4.3. Lessons Learnt

For 2017, GB pedestrian road-crash data show that the results varied. Moreover, the analysis does not clarify reasons causing this variation. Nevertheless, the evidence from this section incite a number of scenarios wherein a pedestrian road-crash casualty could likely happen. The explored method included investigating its applicability (e.g., assessment, verification) by using complex rules as a performance measure [see Sections 6.2 and 6.3]. As a result, on the question of pedestrian general attributes, the male remains nearly twice as much exposed to pedestrian road-crashes than the female. Also, the elderly scored the next highest rates. Moreover, the non-recorded age scored significantly noticeable rates. These findings were unexpected. This because, when it comes to age-related accidents, the data analysis might disqualify. Thus, in the event of an accident, it is recommended undertaking more proactive measures towards age tracking. With respect to the road environment, road-crashes are mostly noticed with dry weather and with permanent speed limits above 30 mph. Additionally, as regards lighting pedestrian road crashes mostly occurred during the day followed by with street light present and lit. Concerning the road infrastructure, the absence of physical facilities, especially pedestrian physical facilities were identified as a key influential factor on road-crashes. Taken together, these provide with relevant new insights. A prototype of this method will be demonstrated in Chapter 8. However, the deployed system indicated two major drawbacks: data inconsistency, lack of relational information. The lack of data consistency restricts its applicability. Lacking relational information between road attributes, as for examples describing road-crashes influenced by street light condition and pedestrian physical facilities as the pelican, might limit the interest of the method to explore undiscovered knowledge. Therefore, improving both the assessment (e.g., road data, ML algorithms) and the data quality consistency creating standardised databases (e.g., GB vs. UK) and quality checks may quite likely apply (Bushema et al., 2013; Jagnoor, 2020).

6.5. Observations

As the undertaken analysis validates its applicability based on the capability of KNIME, and also the data under study, this chapter indicates data quality as key [see Section 6.4.1]. Moreover, the model may be affected from the noisy input data (e.g., misspelling errors, data discrepancies). Subsequently, reducing its reliability when providing results. To mitigate this issue, it is important to consistently quantify the accuracy of the system [see Section 6.3.2]. Nevertheless, KNIME demonstrated its potential to provide objective information. Moreover, the adopted procedure merged data from three distinct contributory factors to pedestrian road-crash, which to date fewer studies dealt with [see Chapter 2]. However, this analysis did not clearly point to other causes affecting pedestrian safety. Moreover, the scope of this study was limited explore the current features of the adopted ML algorithm. Nevertheless, as future work expanding the number and quality of ML algorithms can support the constitution of a standard or a specification (Nashad et al., 2016; Wang et al., 2019). This will be later demonstrated in Chapters 8 and 9.

7. Data Learning Algorithms

7.1. Introduction

This chapter investigates predictive analysis by exploring different ML algorithms previously adopted to improve pedestrian safety. The ML algorithms can best be treated under two main headings: supervised-learning, and unsupervised-learning. Moreover, eligibility criteria to reveal undiscovered patterns, required investigating the impact regarding pedestrian attributes and road elements [see Chapters 4, 5]. The chapter is organised in four main sections. Section 7.1 shows the overall structure of the chapter. Section 7.2 explores discovered knowledge from the ML algorithms both conceptually and in the form of case studies. Section 7.3 explores supervised and unsupervised predictive ML algorithms applied to road data. Section 7.4 discusses the findings.

7.1.1. Overall Approach

Although the applicability of predictive ML algorithms have improved pedestrian safety, a systematic approach on how predictive ML algorithms can contribute is still needed [see Chapter 2]. This chapter explores datamining applied to diverse sets of road data to extract knowledge from predictive ML algorithms. It includes both supervised-learning (Breiman et al.,1984; Zhang, 2019; Premebida et al., 2009; Griselda and Joaquín, 2012; Hariyono and Jo, 2017; Soilán et al., 2018) and unsupervised-learning (Yong et al., 2004; Gosh et al., 2013; Shirkhorshidi et al., 2015; Gu et al., 2017). This is to enhance the implementation of relevant pedestrian safety measures in an urban road environment. One known shortcoming of predictive ML algorithms, their limited capabilities to be transfered to different geographic regions (Serna et al., 2014; Soilán et al., 2018). Another drawback is measuring a reliable accuracy when estimating uncertainty (Abdar et al., 2023; Maryam et al., 2023; Uddin et al., 2023). Nevertheless, studies addressing this challenge have been recently advanced (Gutierrez-Osorio and Pedraza, 2020). Therefore, in addition to conducting further research work, more sophisticated algorithms and techniques for data learning have been previously proposed (Jaber and Rehman, 2020).

7.2. Predictive ML Algorithms

Previous studies have demonstrated various advantages from adopting predictive ML algorithms to explore pedestrian safety elements such as pedestrian general attributes (e.g., gender, age group), the road environment (e.g., weather, speed limits, lighting) and the road infrastructure (e.g., road type, pedestrian facilities) [see Sections 2.2.1, 2.2.3, 2.2.4]. Moreover, when compared with traditional data analysis, predictive ML algorithms have proven to provide pedestrian safety recommendations, more objectively [see Section 2.2.3, Chapter 6]. Figure 7.1 illustrates the two main groups of predictive ML algorithms further explored: supervised vs. unsupervised.

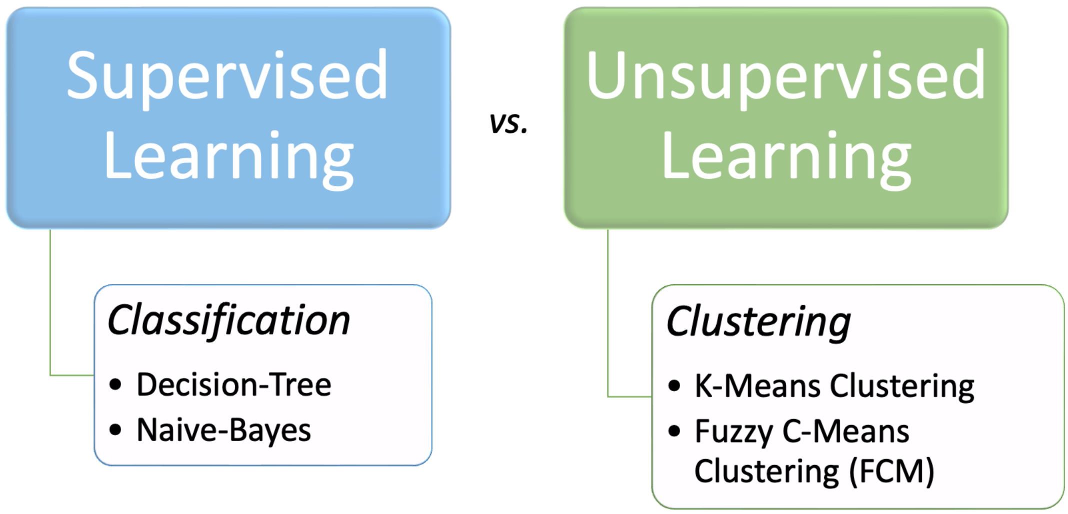


Figure 7.1.: ML algorithms: supervised vs. unsupervised

As per Figure 7.1 supervised-learning classifies road attributes (e.g., gender, age group) and analyses the content of these variables [see Sections 2.5.1, 6.2]. However, typically this type of algorithm demonstrates high dependency on historical events (Burges, 1998; Soilán et al., 2018). In the case of unsupervised-learning algorithms, each road attribute (e.g., gender, age group) corresponds to a cluster that is interpreted according to the combined categories of the most relevant factors (e.g., age group, fatal road-crash = cluster 1, cluster 2, cluster 3) (Ballesteros et al., 2004; Kim et al., 2008; Prato et al., 2012) [see Section 2.5.2]. The advantage here lays on its ability to delineate data patterns independently both with a lower dependency on past data as well as on the number of examined variables (Cottrell and Rousset 1997; Prato et al., 2012; Li et al., 2021). The "Naive Bayes" and “Decision-Tree represented in Figure 6.5, are the supervised predictive ML algorithms explained.

7.2.1. Naive Bayes

Naive Bayes incorporates a family of algorithms based only on a common principle of Bayes classifiers (Silva et al., 2020). The technique uses models that assign class labels, which are variables represented as vectors of feature values that are connected to road elements such as pedestrians and weather. The class labels are drawn from some finite set (Zhang, 2019; Premebida et al., 2009). For example, a pedestrian may be considered classified as a category: gender (e.g., female), age group (e.g., 5-25 years old), height (e.g., 1.65 m). A Naive Bayes classifier, considers each of these features to independently contribute to the probability of determining the pedestrian gender, regardless of any possible correlations between the other variables (e.g., age group, height). One key disadvantage could be the overall accuracy of the algorithms affected by the analysed variables and their associated content (Hariyono and Jo, 2017). To become more accurate on predicting pedestrian general attributes and render Naive Bayes algorithms, the extension of basic motion types and the incorporation of scenes while reducing dependency from historical data, have been previously proposed (Hariyono and Jo, 2017; Soilán et al., 2018) [see Section 2.5.1].

7.2.2. Decision Tree

Decision tree learning is a method commonly used in datamining (Wu et al., 2008). Decision tree builds regression or classification models in the form of a tree structure.

Regression Decision Tree

Regression Decision Tree (RDT) is where the result can be considered a real number. Classification and Regression Tree (CART) modelling was developed by Breiman et al. (1984), and it is technically known as binary recursive partitioning. As a binary decision tree the parent nodes are always split into exactly two child nodes: 0, and 1. This is recursive because the process can be repeated by treating each child node as a parent (Tesema et al., 2005) [see Section 2.5.1]. A few advantages identified from CART models are (Tesema et al., 2005; Li et al., 2020):

i. Easy interpretation of the complex patterns associated with fatality severity.

ii. No requirements for a presumed relationship between the classified attributes (e.g., gender, weather).

iii. Automatic detection of interactions between the attributes through the structure of the tree.

Modelling procedures in the CART: tree building, tree pruning and optimal tree selection (Breiman et al.,1984; Griselda and Joaquín, 2012; Wu et al., 2019). Even though CART methods have successfully validated to model and analyse injury severity (Prato et al., 2010; Griselda and Joaquín, 2012), capturing relevant data to improve accuracy remains an issue (Tesema et al., 2005; Li et al., 2020).

Classification Decision Tree

With respect to Classification Decision Tree (CDT), when a dependent variable is discrete, a classification tree is created [see Sections 2.5, 3.3.3]. The CDT it distinguishes from Classification and Regression Tree (CART) for projecting events from data variables belonging to a class, hierarchically. A study conducted by Serna et al. (2014) used data from two cities: Ohio, USA, and Paris, France. The study was about an automatic and robust CDT approach to detect, segment (i.e., isolated as a single object) and to classify variables from urban road objects belonging to 3D point clouds. In Ohio, the CTD results detected 98% of the objects, segmented 78% and of the adequately segmented objects, classified 82%. On the other hand, the object segmentation in Paris experienced a 15% improvement. Thereby, providing the applicability of CDT modelling (Wu et al., 2008; Serna and Marcotegui, 2014). However, occluded regions, which in 3D means a regions that visible from one side and not visible from the other side, indicated to affect the accuracy of the model from detecting obstacles in real time, and from predicting obstacles ahead in time (e.g., under-segmentation and over-segmentation) (Serna et al., 2014) [see Section 2.5.1]. Turning now to unsupervised ML algorithms, the K-Means Clustering, and the Fuzzy-Clustering-Means (FCM) shown in Figure 6.5 are described next.

7.2.3. K-means Clustering

K-means clustering groups are several observations with common characteristics (e.g., pedestrian behaviour, pedestrian physical facilities) where each observation (e.g., pedestrian gender, zebra) belongs to the cluster with the nearest mean (i.e., cluster centroid) representing the prototype of the cluster (Mohamed et al., 2013). A study conducted between 2002 and 2006, regarded classifying two unique datasets based on pedestrian injury severity. The datasets were categorised into two regions: New York City, USA, and Montreal, Canada. The analysis covered investigating, the pedestrians, the vehicles, and the physical road infrastructure (e.g., geometric design). The findings suggest reducing the exposure of pedestrians to heavy vehicles in a road environment, by for example adding warning signs for the pedestrians in areas with high truck traffic (Aziz et al., 2013). A critical disadvantage is the challenge of relating severity models which generally developed separately. In addition to that, it is important to account that even among individuals in the same age group, the level pedestrian injuries may differ (Milton et al., 2008). To assist this limitation, the deployment of a latent segmentation modelling has been previously proposed (Eluru et al., 2012). According to Bhat et al. (1997), the latent segmentation is based on the order logit distribution and classifies a very high probability as "low risk" and a very low probability as "high risk". Another issue found, was to identify suitable metrics for variables (i.e., datasets) for a particular location when deploying distance-based clustering algorithms. This is because, clustering variables with similar values either in the same or distant clusters, can cause confusion when attempting to connect similar variables. For these cases, limiting the number of datasets (e.g., locations under analysis) have been proposed (Shirkhorshidi et al., 2015).

7.2.4. Fuzzy C-Means

Bezdek (1981) introduced the Fuzzy-Clustering-Means (FCM). The Fuzzy C-Means or FCM is widely applied in agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis and target recognition for feature analysis, and clustering classifier design (Yong et al., 2004). The FCM distinguishes from the K-means since the data can be included in all existing clusters but with varying roles (e.g., pedestrian gender, pedestrian physical fatality) in a range of values rather than being associated to a cluster centroid (e.g., gender, fatal road-crash = cluster 1, cluster 2, cluster 3) (Gosh et al., 2013; Mohamed et al., 2013; Gu et al., 2017). Moreover, a recent study carried out by Gutierrez-Osorio and Pedraza (2020), shows the FCM model recognising the occurrence events at different times and locations operating in real time. Furthermore, the results showed that beyond projecting the future, the approach proposed its expansion to different regions in real time. A study carried out by Li et al. (2021), in Changsha, China, is a good illustration of the FCM applicability in real time. A Fuzzy Cellular Automata (FCA) model was used to evaluate pedestrian safety from conflicts under different proportions of obedient road users (e.g., pedestrians, drivers) and traffic flows. More particularly, the FCA was applied to simulate interactions between pedestrians and vehicles at non-signalised mid-block crosswalks. To collect the data, digital cameras were placed in the observation area (i.e., 4 hours of captured data per site). On the data trajectories of the pedestrians walking speeds crossing in three levels were classified: high, normal, and low. Moreover, a fuzzy inference system was incorporated to describe trends regarding the decision-making process of the drivers in front of a zebra crossing. However, the system failed to measure pedestrian jaywalking behaviour. This is perceived as a key issue as it could have made the road traffic environment more realistic (Li et al., 2021) [see Section 2.5.2].

7.3. Predictive Pedestrian Safety Models: Lessons Learnt

7.3.1. Supervised-Learning Models

In principle, ML algorithms are closely related to regression, classification and ranking [Sections 2.5.1 and 6.3]. As explained in Chapters 4 and 5, to improve precision and accuracy, different models (or rules) are applied depending on the attribute under examination (e.g., age group, type of physical facility). For example, supervised ML algorithms applied to road safety, i.e., supervised-learning models, are used to classify road variables (e.g., gender = 1, 2; road-crash severity = 0, 1, 2) [see Sections 2.5.1 and 7.2]. Moreover, the support-vector machines (SVM) algorithm operates like binary algorithms (Sanz-Casado et al., 2019; Assi et al. 2020). Also, the linear discriminant analysis operates like the logistic regression algorithms (Feliciani et al., 2020). To examine pedestrian road-crashes, supervised-learning algorithms were applied under the following terms:

i. **Pedestrian general attributes:** according to many, severity injury varies according to pedestrian age (Park and Bae, 2020). Therefore, applying different ML algorithms for different age groups is as follows: Logit model for young adolescents [see Section 4.2.3]. Multinominal logit for adults (i.e. neural networks) [see Section 4.2.4]. Logistic regression model for the elderly [see Section 4.2.5].

ii. **Road environment:** for weather, the integer autoregressive model was applied [see Section 5.2.1]. For speed limits, a combination of mathematical formulas was explored [see Sections 5.2.2, 5.2.7]. In the case of lighting, a combination of neural network statistical methods was used [see Section 5.2.3].

iii. **Road infrastructure:** the negative binomial regression was used to examine the influence of intersections in a road environment [see Section 5.2.4]. A regression model was used to investigate the influence of roundabouts [see Section 5.2.5)]. The ordered binary logit was used to analyse the zebra effects [see Section 5.2.6].

7.3.2. Unsupervised-Learning Models

Unsupervised ML algorithms applied to road safety (i.e., unsupervised-learning models), are used to classify road variables (e.g., pedestrian age group fatal road-crash = cluster 1, cluster 2, cluster 3) (Ballesteros et al., 2004; Kim et al., 2008; Prato et al., 2012) [see Section 2.5.2, 7.2]. Regarding FCM algorithms, these seem to enable describing pedestrian road-crashes in real time and ahead in time (Gutierrez-Osorio and Pedraza, 2020). However, no single FCM applied to investigate pedestrian safety through combining road data on pedestrian general attributes (e.g., gender), the road environment (e.g., speed limits) and the road infrastructure (e.g., zebra) was found. One characteristic distinguishing the FCM, is the extension of its potential to different regions and in real time (Aziz et al., 2013; Mohamed et al., 2013; Gutierrez-Osorio and Pedraza, 2020).

7.3.3. Unsupervised ML to improve Pedestrian Safety

Regarding the existing pedestrian safety models, like supervised-learning, the unsupervised-learning demonstrates the ability to analyse data with different models depending on the nature of the attribute under examination. In addition to that, contrary to supervised-learning, unsupervised-learning demonstrates a lower dependency of data from the past. Moreover, the unsupervised-learning validates their ability to describe pedestrian road-crashes in real time and ahead in time (Li et al., 2021). This becomes essential as predicting the pedestrian attributes ahead in time, is highly challenging [see Chapter 4]. Furthermore, in the context of projecting events ahead in time, this study would have been more relevant, if the adopted ML algorithms were more flexible to data with a lower degree of availability and certainty (i.e., data about pedestrians with cognitive disabilities) [see Chapter 2].

7.4. Observations

Despite the extensive advancements on supervised ML models, the FCM indicated to have a high potential to be applied cross-regionally. Moreover, compared with supervised-learning, the FCM proves to require a lower amount of data to operate. Furthermore, without being explicitly programmed for a specific purpose, this data mining technique allows computers to learn, interpret and discover new knowledge from both small and large sets of data (Aziz et al., 2013; Mohamed et al., 2013; Gutierrez-Osorio and Pedraza, 2020). To improve the FCM precision and accuracy, so that predictions can be trusted for in a real road environment (e.g., decision-making, pedestrians), prior to incorporating the model in real-time data, pre-assessing data and comprehensively defining its features is vital. This is because, beyond data inconsistency or incompletion, cyber security may also intervene. Therefore, the adoption of more sophisticated pedestrian safety models to not only detect incomplete data, but to correct unforeseen failures is also suggested [see Section 2.2]. This chapter examined several previously adopted ML predictive models to examine pedestrian safety. The next chapter synthesises in the form of a prototype, a data-driven analysis using the FCM to investigate influential factors on pedestrian road-crash casualty causation.

8. Prototype Model

8.1. Introduction

To date a systematic investigation combining the pedestrian general attributes phenomena with the road elements have not been dealt with much detail (De Oxley et al., 2005; Bungum et al., 2005; Chen and Persaud, 2012; Treno et al., 2014; Papadimitriou et al., 2015; Jones and Anciaes, 2017; Hezaveh and Cherry, 2018). Moreover, the ability to accurately measure and assess different road attributes jointly remains a challenge (Li et al., 2021) [see Chapters 4, 5; Section 7.2.4]. This chapter explores the applicability of the Fuzzy-Clustering-Means (FCM), which is an unsupervised ML algorithm (Gutierrez-Osorio and Pedraza, 2020; Li et al., 2021). The FCM prototype was deployed to systematically investigate pedestrian general attributes and the influence of road elements on pedestrian safety. This study relied on spatial analysis using data from a single year [see Chapters 2, 6, and 7]. According to Chapters 4 and 5, most pedestrian predictive models focus on singular factors affecting pedestrian safety. Moreover, such approaches have, however, failed to address examining multiple road attributes in a singular analysis. Therefore, this study is considered to offer relevant contributions to modeling development. This is because the prototype proved its potential to simultaneously examine multiple factors. This means that, should the model of choice have the capability to predict attributes ahead of time, it can more efficiently offer new knowledge on road-crash trends. Since the prototype proved its applicability, further research on how to integrate the most applicable models in a single examination, particularly considering that the nature of the data is distinct and that it varies, is highly recommended. Moreover, such approaches, however, have failed to address examining multiple road attributes on a singular analysis. Therefore, this is study considered to offer relevant contributions to the modelling development. This is because, the prototype proved its potential to simultaneously examine multiple factors. This means that, should the model of choice have capabilities to predict attributes ahead in time, it can more efficiently offer new knowledge on road-crash trends. Furthermore, as the prototype proved is applicability, further research on how to integrate on a single examination the most applicable models, particularly considering that the nature of the data is distinct and that it varies, is highly recommended. A flow chart of the prototype model deployed is shown in Figure 8.1.

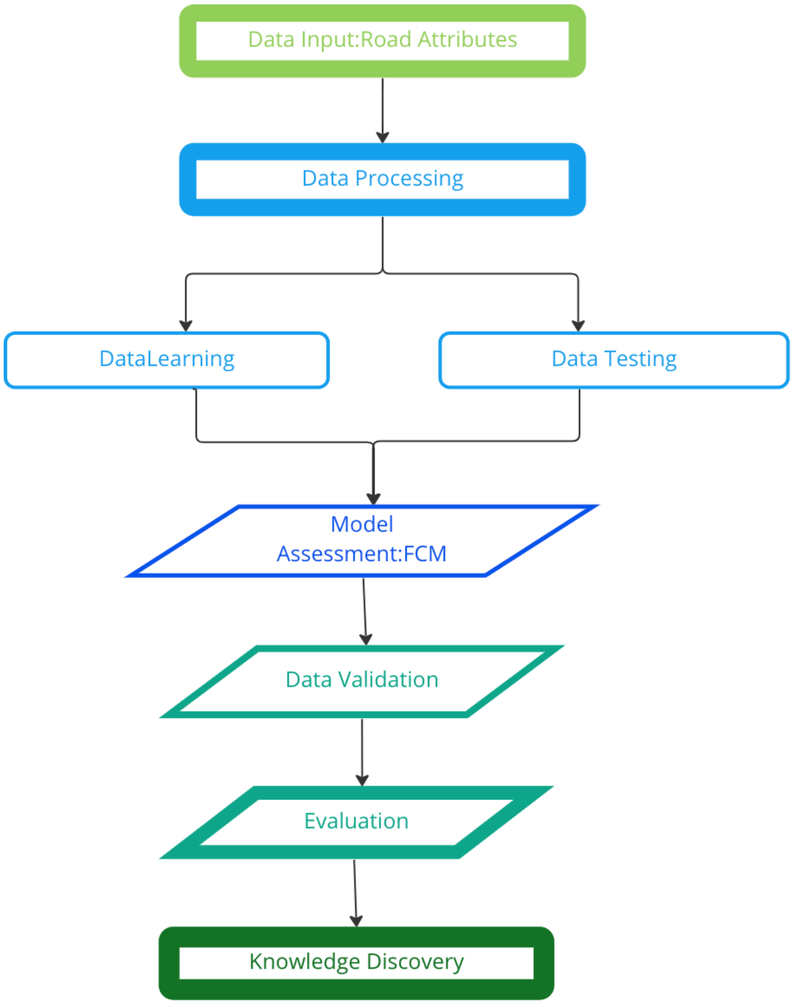


Figure 8.1.: Prototype model diagram

The choice of the FCM (i.e., prototype) presented in Figure 8.1 to examine the road data, was for its demonstrated suitability as a pedestrian safety model on different urban road environments (Aziz et al., 2013; Mohamed et al., 2013). In addition to that, for is potential of being applicable to be used in real-time environments (Li et al., 2021) [see Section 7.3.3]. A more detailed description on how the prototype model works can be found in Appendices A.3.1 and A.3.2. The prototype was deployed to systematically investigate the influence on pedestrian safety from combining a few pedestrian general attributes and road elements. Some fundamental contributions of the modelling development which need to be addressed, are the consistency of discovered knowledge to improve pedestrian safety. This knowledge can be acquired from combining different data such as the pedestrians and the road infrastructure characteristics. In addition to that, compared to previous approaches, it enables using less data (i.e., one year) while enabling the possibility of incorporating different algorithms in the same analysis (Yannis and Karlaftis, 2010; Torbic et al., 2010) [see Section 3.2.2]. This chapter covers an in-depth examination on the impact of pedestrian general attributes (i.e., gender, age group), the road environment (i.e., weather, speed limits, lighting) and the road infrastructure (i.e., intersection type, pedestrian facilities type) using the prototype to describe pedestrian safety patterns. To achieve such, a systematic descriptive analysis using the above attributes was verified and validated. The chapter is organised in four main sections. Section 8.1 shows the overall structure of the chapter. Section 8.2 presents the data verification analysis procedure to assess the model applicability using the FCM demonstrator on two distinct sets of data: GB and London. This was experimented, to prove its suitability in different regions [see Chapters 3, 6]. In Section 8.3, the FCM algorithm is applied on GB road data to discover new knowledge. Section 8.4 discusses the lessons learnt.

8.2. Data Analysis Verification

In general this study describes the main types of data analysis which are conventional and ML [see Section 3.2.1]. The conventional analysis was used to deploy a temporal analysis. This means gathering pedestrian road-crash casualty data during an 11-year period (i.e., 2009-2019) (DfT, 2017; DfT, 2018; DfT, 2019). This approach will be presented later on in Chapter 10. As regards ML analysis the adopted FCM was applied on a spatial analysis respectivelly using data from London and GB (DfT, 2017). The ML algorithms are generally described in two modes: descriptor and predictor (Buscema, 2013) [see Section 2.5; Section 3.2.1]. The FCM applicability was verified in Section 8.2.1 and further expanded in Section 8.3. Sections 8.2.1 and 8.2.2, explored data in the form of a demonstrator in two regions: GB and London. This allowed for extracting knowledge using:

i. A sample of actual road data from 2017.

ii. The sub-mode descriptor to train the data.

iii. The sub-mode predictor to test the data [see Chapter 3].

Additionally, the same source of data (i.e., STATS19) was used to separately extract data from the two explored regions. For the ML data analysis, the statistical data was manually incorporated on KNIME, and for illustration added to a visualisation tool called Google Earth Education (KNIME; DfT, 2017; Google Earth Education, 2021). The validation of the prototype applied the FCM sub-modes (i.e, descriptor and predictor) to two distinct sets of data regarding pedestrian road-crash casualties in 2017: GB and London [see Sections 8.2.2 and 8.2.4]. Additionally, the results of the sub-modes using actual data from 2017 to project pedestrian road-crash casualty trends, were compared with actual data from 2018.

8.2.1. Data Statistics Verification: GB

PMSLo Figure 8.2.1 displays GB pedestrian road-crash casualties in the year of 2017, by road type [1].

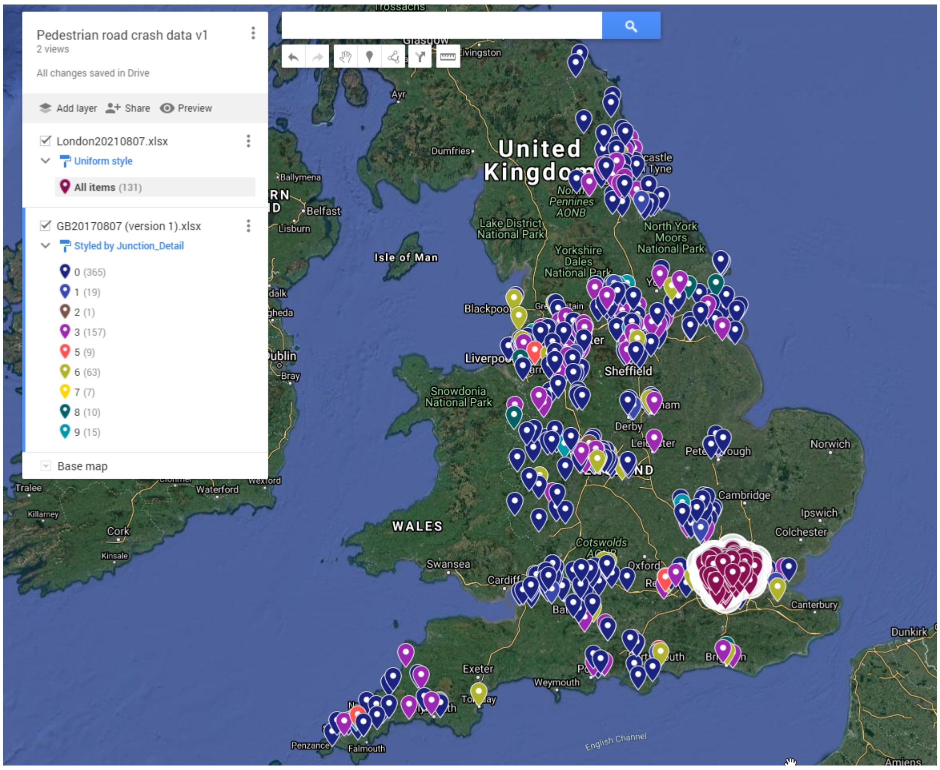


Figure 8.2.: GB pedestrian road-crash casualties

Source: (DfT, 2017; Google Earth Education)

Figure 8.2.1 shows most accidents (i.e., 365) occurring when road physical facilities are absent (i.e., 0 - No Junction).

[1] 0- No Junction(NJ); 1- T staggered(TS); 2- Roundabouts(R); 3- Mini-roundabouts(MR); 4- Slip road(SL); 5-Cross roads(CR); 6- Private drive(PD); 7- Multiple Junction(MJ); 8- Other Junction(OJ); 9- Other (O)

8.2.2. FCM Verification Demonstrator: GB

The applicability of the FCM demonstrator was tested on actual GB road data from 2017 (DfT, 2017). The FCM verification demonstrator splits the data into two sub-modes: descriptor and predictor. The descriptor is used to train 50% of data, and the predictor is used to test 50% [see Section 6.3.3]. In relation to the accuracy, a value greater than 70% indicates an acceptable prediction (Mendoza et al., 2023; Noor et al., 2023). Table 8.1 shows the extracted knowledge from the FCM (i.e., descriptor, and predictor) regarding the total GB pedestrian road-crash casualties.

Table 8.1.: Data verification: road fatalities per accident

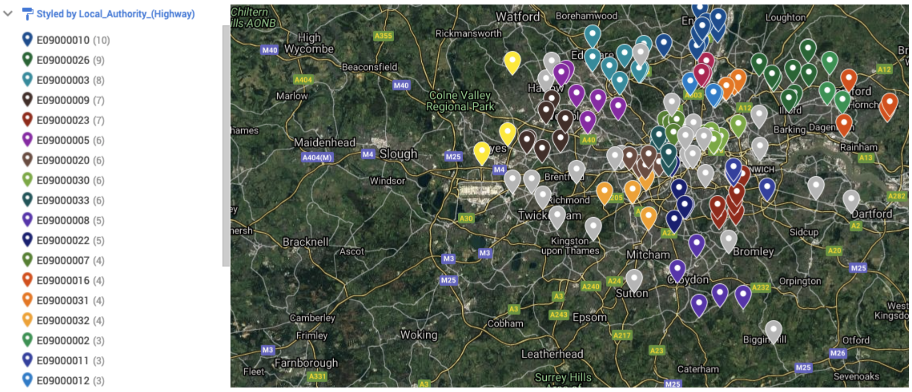
GB number of pedestrian fatalities: 470	Descriptor (50% of the total data)/Accuracy %	Predictor (50% of the total data)/Accuracy %
Data Trained: KNIME	235/ 97.03%	235/ 96.81%
1	69.18/N/A	73.29/N/A
2	15.07/N/A	20.22/N/A
3	7.19/N/A	2.17/N/A
4	1.37/N/A	N/A/N/A
5	1.71/N/A	1.81/N/A
6	2.05/N/A	N/A/N/A
7	N/A/N/A	2.53/N/A
10	3.42/N/A	N/A/N/A

As presented in Table 8.1, the descriptor and the predictor sub-modes applied the rules respectively. The FCM descriptor is represented by the data set on the right side and the FCM predictor is represented by the dataset on the left side). Regarding the accuracy other than the total number of fatalities (470), no information on the number of fatalities per accident was found. Therefore, in comparison to actual data from 2018 the accuracy was only possible to measure for the descriptor (97.03%) and the predictor(98.81%) (Assi et al., 2020). In terms of the FCM descriptor, GB road date indicates an overall 0.1% decrease in road fatalities (i.e., from 235 to 234.98). In addition to that, it indicates a 69.18% chance of one pedestrian fatality and a 1.71% chance of five fatalities. Moreover, interestingly the descriptor also shows a 3.42% chance of up to 10 fatalities. The latter, seems rather intricate. The system, however, was unable to describe the number of accidents that could result in seven fatalities. Moreover, it does not present any information on the number of accidents that could result in eight or nine fatalities. On the other hand, the overall findings for the FCM predictor indicate a 0.2% increase in fatalities (from 235 to 235.53). Additionally, it indicates a result of one accident, a 73.29% chance of one fatality and a 1.81% chance of five fatalities. In this case, contrary to the FCM descriptor the FCM predictor does not present any data (i.e., N/A) on a chance of up to four, six, and ten fatalities accordingly. Nonetheless, it indicates a 2.53% chance of seven fatalities, whereas the FCM descriptor yielded no results (i.e., N/A). The actual trends between 2017 and 2018 demonstrate a decreasing tendency by 2.98% (from 470 to 456) (DfT, 2017; DfT, 2018). In overall the trends align with the FCM descriptor and on the contrary are misaligned with predictor. Also, in terms of data accuracy both the FCM descriptor and the FCM predictor fail to precisely calculate the percentage of decreasing propensity between 2.5% and 3%. Moreover, regarding the number of fatalities per accident identified for 2017, for 2018 no detailed information was found. This is a key issue that limits the system accuracy validation: a 69,18% chance of one fatality, with no data from 2018 to be compared with (i.e., 69,18/N/A). One other issue found was the difference in numbers of the total number of GB pedestrian fatalities by 34, between the officially reported data (i.e., 456) and the examined database (i.e., 422) (DfT, 2017; DfT, 2018; STATS19). These results suggest a more detailed examination which will be carried out in Section 8.3. Nevertheless, the FCM transferable capabilities to different regions, will be introduced on a similar verification analysis, for the London region in Sections 8.2.3 and 8.2.4.

8.2.3. Data Statistics Verification: London

Figure 8.3 shows the London Region pedestrian road-crash casualties in the year of 2017.

(a) Pedestrian road-crash casualty rates in 2017: London region



(b) London region pedestrian road-crash casualty rates in 2017 by local authority

Figure 8.3.: London Region pedestrian road-crash casualty rates in 2017 Source: (DfT, 2017; Google Earth Education)

Figure 8.3a summarises the total accident rates (i.e., 131) for the London Region in 2017. Figure 8.3b classifies the London Region accident rates in 2017, by local authority (STATS19) [see Appendix A.2]. In overall, in 2017, data show Enfield with most road-crashes involving pedestrians (i.e., 10). These results are directly aligned with the reported data from Transport from London (TfL, 2018).

8.2.4. FCM Verification Demonstrator: London

Similar to Section 8.2.2, the FCM verification demonstrator was applied on actual road data from 2017 for London (DfT, 2017). Table 8.2 illustrates the extracted knowledge from the FCM on the total pedestrian road-crash casualties in London.

Table 8.2.: Road-crash casualty data verification: scenario vs. KNIME

London number of pedestrian fatalities: 131	Descriptor (50% of the total data)/ Accuracy %	Predictor (50% of the total data)/ Accuracy %
Data Training: KNIME	65.5/ 85.11%	65.5/ 85.24%
1	63.27/N/A	74.72/N/A
2	12.24/N/A	18.6/N/A
5	10.6/N/A	6.98/N/A
7	14.29/N/A	N/A/N/A

As presented in Table 8.2 the FCM descriptor is represented as the first dataset and the FCM predictor as the second dataset [see Section 6.3.3]. Regarding the accuracy other than the total number of fatalities (131), no information on the number of fatalities per accident was found. Therefore, in comparison to actual data from 2018 the accuracy was only possible to compare the results with the total number of records. In terms of accuracy, the descriptor reached 85.11% and the predictor 85.24% (Assi et al., 2020). More in detail, in relation actual data on pedestrian fatalities in 2018, for London the FCM descriptor indicates an overall 0.4% increase in road fatalities (i.e., from 65.5 to 65.8). It also shows a 63.27% chance of one fatality, a 12.24% chance of two fatalities, a 10.6% chance of five fatalities, and a 14.29% chance of seven fatalities. On the other hand, for the FCM predictor, the overall findings indicate a 0.3% increase in fatalities (from 65.5 to 65.7). Additionally, the predictor shows a 74.72% chance of one fatality, an 18.6% chance of two fatalities, a 6.98% chance of five fatalities. However, it does not show any findings on a chance of up to seven fatalities (i.e., N/A). Moreover, it does not present any data on the number of accidents that could result in three, four and six fatalities. Regarding the actual pedestrian fatality rates between 2017 and 2018 show a 15.5% decrease (i.e., from 131 to 112) (TfL, 2018). Moreover, for the most risky region of London for road-crash fatalities (i.e., Enfield) it shows a 20 % decreasing tendency (from 10 to 8) (TfL, 2018). The findings in overall are contrary to both the FCM descriptor and predictor overall results. Also, similar to the FCM verification demonstrator in Section 8.2.1, the system fails to calculate road-crashes percentage propensity. Moreover, for 2018 no evidence discerning the number of fatalities per accident was found (TfL, 2018). Similar to Section 8.2.2, for the London region, road data between the officially reported data (112) and the examined database (57), shows a difference on total number of pedestrian fatalities by 55 (TfL, 2018). In both the analysed regions this is rather critical to: describe results, validate the findings, and project results ahead of time. As previously explained in Chapter 6, these results seem to happen due to data inconsistency and lack of relevant relational information and also undefined data harmonisation procedures [see Sections 3.3.3, 6.4.2, and 6.4.3]. In summary, the examination was adopted to verify the number of pedestrian road-crash fatalities, by road type (i.e., for GB), and by local authority (i.e., for London) as single elements. The FCM Demonstrator, applied both the descriptor and the predictor sub-modes. However, the results from the predictor, highly differ from reality [see Section 8.2]. Also, when examining the same dataset, the descriptor failed to indicate results (i.e., N/A) in one out of 12 and the predictor failed in four out of 12. The latter indicates to be a less consistent. Moreover, the descriptor ranked the highest accuracy rate (97.3%). Therefore, in this study, the FCM descriptor (i.e., train 80%) was applied to GB road-crash casualty data from 2017 to carry out the systematic investigation. In addition to that, the results were then contrasted with actual road data from 2018 (i.e., statistics) (DfT, 2018). Furthermore, as some variations may be related to system sensitivity response, a further explanation will be carried out later in Chapter 9. Nevertheless, the exercise indicated to be applicable to both GB and London regions [see Sections 8.2.2 and 8.2.4].

8.3. Knowledge Discovery: GB Pedestrian Road-Crash Data

The knowledge discovery activity investigates a combined number of elements previously investigated in Chapters 4 and 5. More in detail, it combines exploring road data on pedestrian general attributes by gender and age group, on road environment by weather, speed limits and lighting, and on road infrastructure by road type and pedestrian physical facilities. However, Section 8.2 indicates that when using the FCM verification demonstrator on both the GB and London regions, the system indicates a significant discrepancy between the projected data and the actual data. Moreover, the data of the two sub-modes (i.e., trained and tested) in both Sections 8.2.2 and 8.2.4, proves that the information projected with failure is significantly higher for the FCM predictor than for the FCM verification demonstrator. Subsequently, the prototype only examined the results of the FCM descriptor trained on an 80% sample ratio of the actual GB road data from 2017 (i.e., statistics). Afterwards, the results were compared with GB road data from 2018 (i.e., statistics) from the official sources (DfT, 2017; DfT, 2018). To measure the accuracy, unlike for the FCM demonstrator, which was applied to a relatively small dataset in Sections 8.2.2 and 8.2.3, the performance was assessed using a standard metrics widely used to evaluate the performance of ML models, called margin of error (MOE) (Tanur, 2011; Uddin et al., 2023). The MOE is a statistical measure used to verify the FCM descriptor; its accuracy measures its reliability with sample data (i.e., 470 fatalities). Essentially, it is the number of random sampled error in the results to reflect the number of uncertainty of the FCM descriptor. The larger the MOE, the less the results described by the FCM descriptor (used to project data ahead in time) would reflect the actual results. For the analysed dataset a MOE rate of ca. 3% suggests a quite accurate system (James et al., 2015). The statistical calculation of the MOE is given by,

$$MOE(\gamma) = Z(\gamma) \times \sqrt{\frac{\sigma^2}{n}} \tag{8.1}$$

where z-scores is denoted by Z(γ), the standard error is given by

$$\sqrt{\frac{\sigma^2}{n}}, \tag{8.2}$$

and n is the number of pedestrian casualties.

(Tanur et al., 2014)

In addition to MOE, the z-score was used as a performance metric. The z-score or standard score, is the number of standard deviations above or below of the average pedestrian casualties. According to some, a optimal z-score falls between -2.0 and +2.0 (Sakai et al., 2015).

8.3.1. Pedestrian General Attributes

Road-Crash by Gender: Statistics vs. Descriptor

Figure 8.4 compares the findings from 2017 statistics applied on the FCM descriptor with actual data from 2018 on pedestrian general attributes by gender.

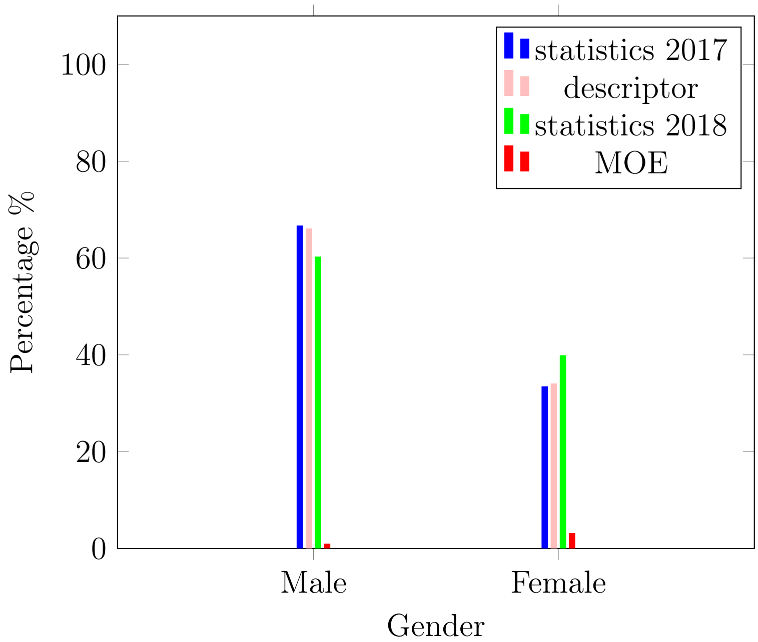


Figure 8.4.: Road-crash by gender: statistics (2017, 2018) vs. descriptor

Table 8.3.: Road-crash by gender: descriptor vs. margin of error

Pedestrian fatalities: 470	Gender: %	MOE: %	Z-Score
Male/Female	66.01/ 33.99	0.88/3.09	0.17/ -0.30

As can be seen in Figure 8.4 and Table 8.3, for GB the descriptor shows a 66.01% chance of fatality for males and a 33.99 % chance for females. Compared with the statistics data from 2018, the descriptor indication aligns with results from the 2018 (DfT, 2017; DfT, 2018). The calculated MOE and the z-score metrics shown in Table 8.3, are within an acceptable performance range [see Section 8.3].

Road-Crash by Age Group: Statistics vs. Descriptor

Figure 8.6, compares the findings between the statistics from 2017 applied on the FCM descriptor with actual data from 2018 on pedestrian general attributes by age group[2].

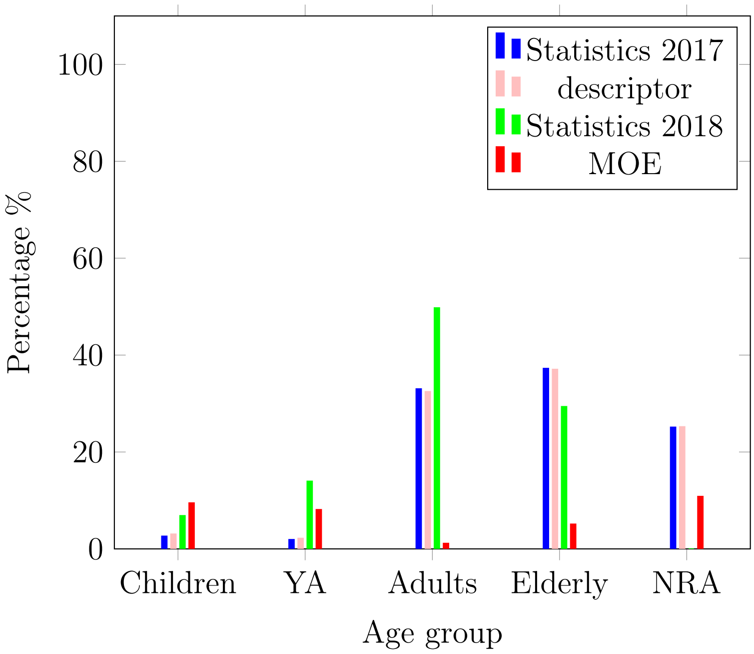


Figure 8.5.: Road-crash by age group: statistics (2017, 2018) vs. descriptor

Table 8.4.: Road-crash by age group: descriptor vs. margin of error

Pedestrian fatalities: 470	Age Groups: %	MOE: %	Z-Score
Children; Young	3.07; 2.19	9.50; 8.12	-2.69; -1.03
Adults; Elderly	32.47; 37.07	1.15; 5.12	-0.16; -0.56
Non recorded ages	25.22	10.84	-1.08

As shown in Figure 8.6 and Table 8.4 the descriptor shows a 37.07% chance of fatality for elderly, a 32.47% chance of fatality for adults, a 3.07% chance of fatality for children, a 2.19% chance of fatality for young adolescents. Compared

[2] Children (0-14); Young adolescents (YA) (15-24); Adults (25-64), Elderly (65+); Non recorded ages (NRA)

with the statistics data from 2018, a significant increasing tendency for children, young adolescents, and adults is present. On the contrary, a decreasing tendency for the elderly is noticed. Except for the adults, the descriptor indication aligns with the statistics from 2018. As per Table 8.4, the calculated the MOE presents an acceptable performance only for adults. For the z-score metrics it only presents an acceptable performance for children [see Section 8.3]. However, the large number of unrecorded age rates (approximately 25.22%) may significantly affect to measure the metrics of performance (DfT, 2017; DfT, 2018).

8.3.2. Road Environment

Road-Crash by Weather: Statistics vs. Descriptor

Figure 8.8 compares the findings between statistics from 2017 applied on the FCM descriptor with actual data from 2018 on road environment by weather[3].

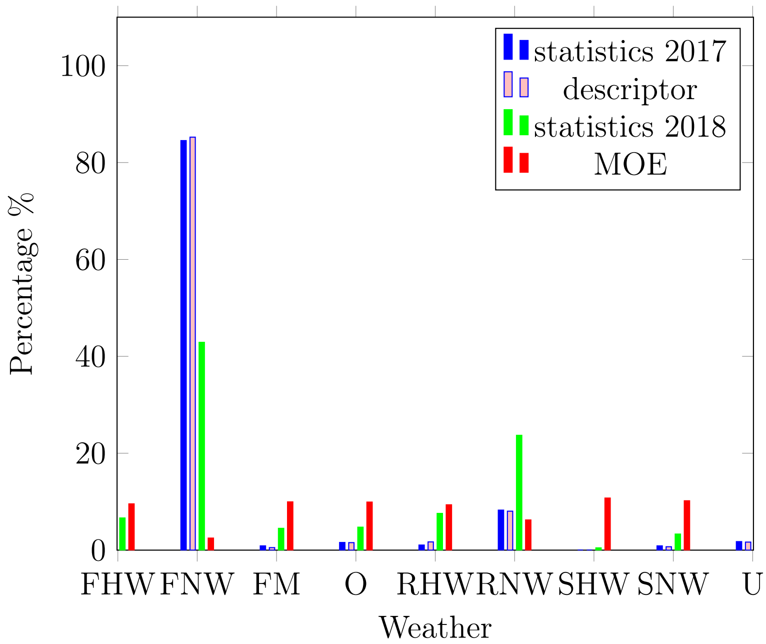


Figure 8.6.: Road-crash by weather: statistics (2017, 2018) vs. descriptor

Table 8.5.: Road-crash by weather: descriptor vs. margin of error

Pedestrian fatalities: 470	Weather: %	MOE: %	Z-Score
FHW; FNW	0.6; 85.24	9.55; 2.49	-1.52; -0.28
FM; Other	0.52; 1.54	9.96; 9.92	-1.24; -1.25
RHW; RNW	1.7; 8.04	9,36; 6,23	-1,10; -0.74
SHW; SNW	N/A; 0.7	10.75; 10.19	-1.28; -1.29
Unknown	1.66	9.64	-1.24

As can be seen in Figure 8.8 and Table 8.5 the descriptor indicates FHW with a

[3]Fine-high winds (FHW); Fine-no winds (FNW); Fog or mist (FM); Other (O); Raining-high winds (RHW); Raining-no winds (RNW); Snowing-high winds (SHW); Snowing-no winds (SNW); Unknown (U)

0.66% chance of fatality, FNW with a 85.24% chance of fatality, FM with a 0.52% chance of fatality, Other with a 1.54% chance of fatality, raining with a 1.7% chance of RHW with a 8.04% chance of fatality, SHW with a 0.7% chance of fatality and Unknown with a 1.66% chance of fatality. Compared with the statistics data from 2018, the descriptor indicates opposite results to the actual data (DfT, 2017; DfT, 2018). As per Table 8.5, the calculated the MOE presents the best performance rate only for FNW, which is where the highest chance of fatality is noticed. The z-score metrics presents an acceptable performance for all the types of weather [see Section 8.3].

Road-Crash by Speed Limits: Statistics vs. Descriptor

Figure 8.7 compares the findings between statistics from 2017 applied on the FCM descriptor with actual data from 2018 on road environment by the speed limits[4].

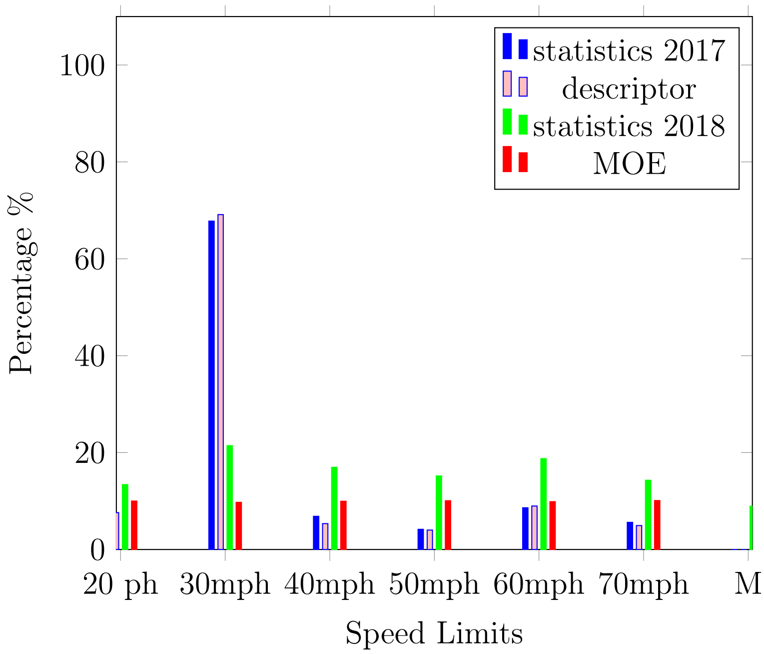


Figure 8.7.: Road-crash by speed limits: statistics (2017, 2018) vs. descriptor

Table 8.6.: Road-crash by speed limits: descriptor vs. margin of error

Pedestrian fatalities: 470	Speed Limits: %	MOE: %	Z-Score
20 mph; 30 mph	7,61; 69,12	10,15; 9,73	-2,24; -1,18
40 mph; 50 mph	5,34; 4,01	9,96; 10,06	-1,20; -1,19
60 mph; 70 mph	8,97; 4,94	9,87; 10,10	-1,21; -1,20
Motorway	0	10.38	-1.28

In Figure 8.7 and Table 8.6 the descriptor shows a 7.61% chance of fatality for 20 mph, a 69.12% chance of fatality for 30 mph, a 5.34% chance of fatality for 40 mph, a 4.01% chance of fatality for 50 mph, a 8.97% chance of fatality for 60 mph, a 4.97% chance of fatality for 70 mph (DfT, 2017; DfT, 2018; DfT, 2019). Compared

[4] 1-20 mph, 21-30 mph, 31-40 mph, 41-50 mph, 51-60 mph, 61-70 mph, Motorway (M)

with the statistics data from 2018, except for the alignment with 20 mph and the 60 mph, the descriptor indicates opposite tendency results from the actual data (DfT, 2017; DfT, 2018). As per Table 8.6, the calculated the MOE does not present an acceptable performance for any of the speed limits. The z-score metrics presents an acceptable performance except for 20 mph [see Section 8.3].

Road-Crash Lighting: Statistics vs. Descriptor

Figure 8.8 compares the findings between statistics from 2017 applied on the FCM descriptor with actual data from 2018 on road environment by the lighting condition[5].

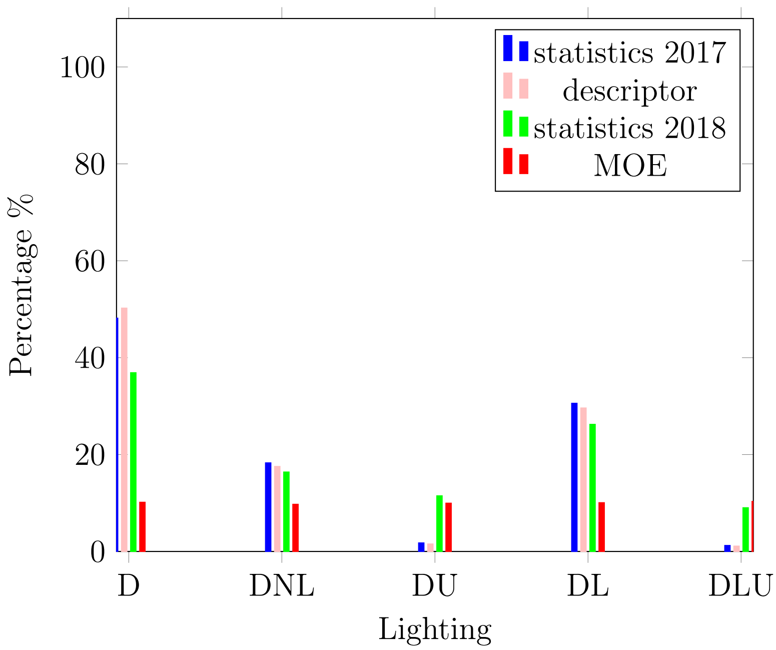


Figure 8.8.: Road-crash by lighting: statistics (2017, 2018) vs. descriptor

Figure 8.8 and Table 8.6 detail the descriptor with a 50.22% chance of fatality at daylight, a 29.61% chance of fatality at night with the lights unlit, a 17.54% chance

[5]Daylight (D); Dark no lighting (DNL); Dark unlit (DU); Dark lit (DL);Dark light unknown (DLU)

Table 8.7.: Road-crash by Lighting: descriptor vs. margin of error

Pedestrian fatalities: 470	Lighting: %	MOE: %	Z-Score
D; DNL	50,22; 17,54	8,76; 9,92	-9,52; -5,27
DU; DL	1,54; 29,61	10,19; 9,36	-4,76; -4,87
DLU	1,54	10.33	-5.30

of fatality at night with no lighting, a 1.54% chance of fatality at night with the lights lit, a 1.1% chance of fatality at night with the lights unknown. Compared with the statistics data from 2018, the descriptor indication aligns with the results on Dark no lighting and Dark lit. On the contrary, the alignment at Daylight, Dark unlit, and Dark light unknown disagrees with the actual data (DfT, 2017; DfT, 2018). As shown in Table 8.7, neither the calculated the MOE nor the z-score present an acceptable performance for any of the assessed lighting condition [see Section 8.3].

8.3.3. Road Infrastructure

Road-Crash by Road Type: Statistics vs. Descriptor

Figure 8.9 compares the findings between statistics from 2017 applied on the FCM descriptor with actual data from 2018 on road infrastructure by road type[6].

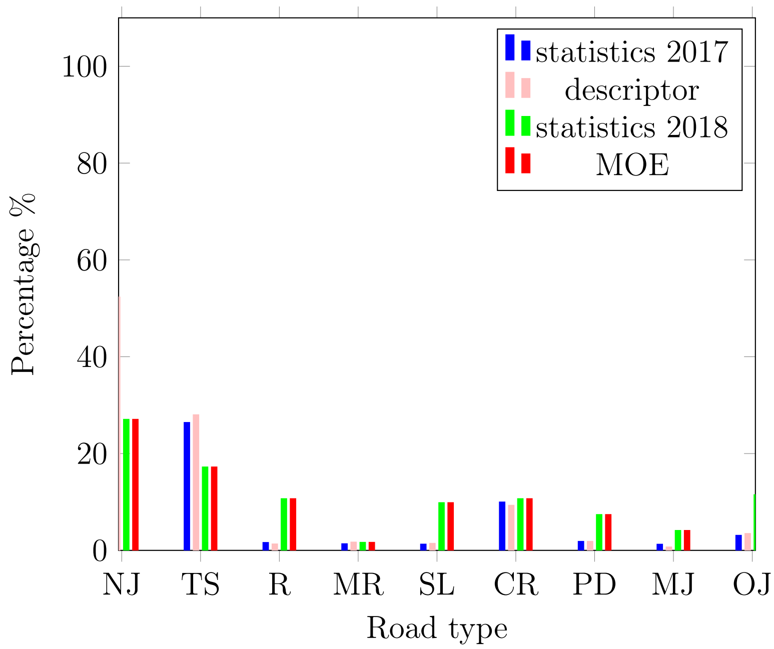


Figure 8.9.: Road-crash by road type: statistics (2017, 2018) vs. descriptor

Table 8.8.: Road-crash by road type: descriptor vs. margin of error

Pedestrian fatalities: 470	Road Type: %	MOE: %	Z-Score
No junction; T staggered	52,32; 27,99	9,32; 9,87	-2,38; -1,41
Roundabouts; Mini-Roundabouts	1,31; 1,69	10,24; 10,75	-1,37; -1,44
Slip road; Crossroads	1,43; 9,29	10,29; 10,24	-1,22; -1,22
Private drive; More than 4 arms	1,84; 0,68	10,42; 10,61	-1,24; -1,28
Other Junction	3.46	10.19	-1.23

[6] No junction (NJ); T staggered (TS); Roundabouts (R); Mini-roundabouts (MR); Slip road (SL); Cross roads (CR); Private drive (PD); Multiple Junction (MJ); Other Junction (OJ)

Figure 8.9 details the descriptor with a 52.32% chance of fatality at NJ, a 27.99% chance of fatality at TS, a 9.29% chance of fatality at CR, a 3.46% chance of fatality at OJ, a 1.84% chance of fatality at PD, a 1.69% chance of fatality at MR, a 1.43% chance of fatality at SL, a 1.31% chance of fatality at roundabouts, a 0.68% chance of fatality at MJ (DfT, 2017; DfT, 2018; DfT, 2019). Compared with the statistics data for 2018, except for alignment at other junctions, the descriptor indicates opposite tendency results from the actual data (DfT, 2017; DfT, 2018). In Table 8.8, the calculated MOE does not present any metric with an acceptable performance for any road type. The z-score metrics present an acceptable performance except for at no junction [see Section 8.3].

Road-Crash by Pedestrian facilities: Statistics vs. Descriptor

Figure 8.10, displays the findings between the statistics from 2017 and the FCM descriptor on road infrastructure by type of pedestrian facilities[7].

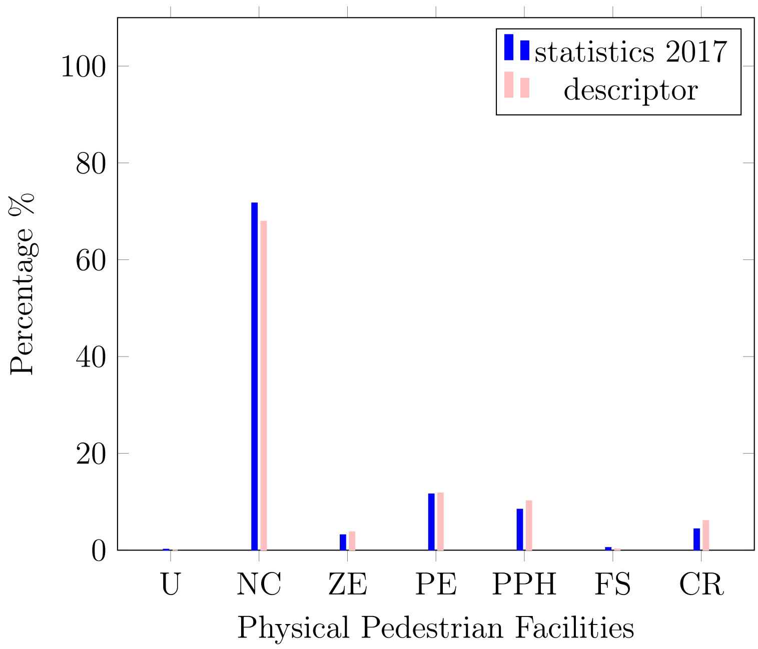


Figure 8.10.: Road-Crash by Pedestrian facilities: statistics (2017) vs. descriptor

In Figure 8.10 the descriptor details a 67.92% chance of fatality at no pedestrian facilities, a 11.81% chance of fatality at pelican, a 10.17% chance of fatality pedestrian phase, a 6.1% chance of fatality at central refuge, a 3.77% chance of fatality at zebra, a 0.23 % chance of fatality at footbridge or subway. However, statistics from 2018 were not found. Compared with the statistics data from 2017, a slight decreasing tendency at no physical facilities is described. Also, a slight increasing tendency at the remaining pedestrian facilities type (DfT, 2017; DfT, 2018).

[7]Unknown (U); None crossing within 50 m (NC); Zebra (ZE); Pelican (PE); Pedestrian Phase (PPH); Footbridge or subway (FS); Central refuge (CR)

8.4. Lessons Learnt

In this study the FCM demonstrator was verified in the descriptor, and predictor sub-modes. An acceptable accuracy rate should score above 70% [see Section 8.2.2]. The FCM demonstrator scored 85.11% as the minimum for the London dataset, and 97.03% as the maximum for the GB dataset. Especially the London dataset, the results showed discrepancies between the predictions and the actual data. Therefore, the FCM prototype used the best accuracy rate, i.e., the descriptor sub-mode to explore GB road data. In addition to that, to assess each of the examined datasets in relation to actual data, additional metrics (i.e., MOE, z-score) were applied [see Section 8.3]. Figure 8.11 summarises the findings of the FCM prototype regarding the accuracy rates compared with statistical data from 2018.

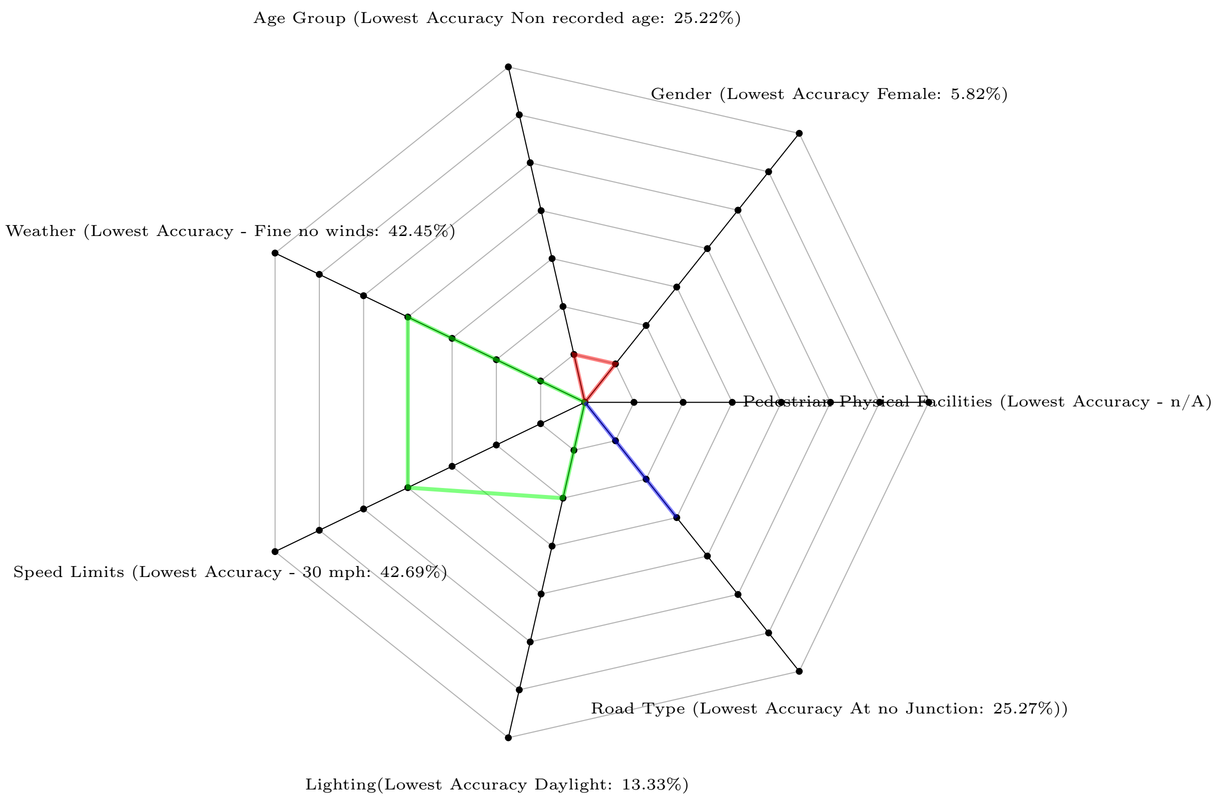


Figure 8.11.: The seven examined influential factors affecting pedestrian safety

Accuracy by

— Gender and Age Group

— Weather, Speed Limits and Lighting

— Road Type and Pedestrian Physical Facilities

The illustrated scenario in Figure 8.11 indicates that the prototype describes the gender with the highest level of accuracy (i.e., female: 5.82%). The calculated MOE and z-score agree [see Sections 8.3 and 8.3.1]. The results on the gender described a tendency to a chance of fatality decreasing both for males and increasing for females. These results align with statistics from 2018 which becomes rather encouraging. With respect to the age group, except for the adults, the results described a slight increasing tendency to a chance of fatality. These findings mostly align with the statistics from 2018. However, it is important to highlight the large percentage (ca. 25.22%) of NRA as a key issue. This is because, the NRA were described with a slight increasing tendency when the non-recorded age rates for 2018 data were not found. These findings seem critical. Therefore, a further investigation is proposed as the lack of data might not only affect the results by age group, but the entire validity of the analysis [see Section 8.2.2]. In terms of road environment, it indicates the lowest level of accuracy for speed limits (i.e., 30 mph: 42.69%) and the weather (i.e., fine no winds: 42.45%). About the performance metrics for the speed limits the MOE agrees whereas the Z-score disagrees [see Sections 8.3 and 8.3.2]. On the other hand, for the weather both the calculated MOE and z-score agree [see Sections 8.3 and 8.3.2]. As regards weather, it described a tendency to a chance of fatality mostly to happen with fine dry weather. Here, the results indicated an opposite tendency to the statistics from 2018, hence sharing fewer promising results. Regarding the speed limits, the results described a tendency to a chance of fatality indicate to mostly happen at 30 mph. Except for the alignment for 20 mph and the 60 mph, the results indicated opposite tendency results from the statistics from 2018. Concerning lighting, the results described a tendency of a chance of fatality to mostly happen at daylight. The lighting condition results mostly indicates alignment for a chance of a fatality at night. Regarding the road infrastructure, these described a tendency to a chance of fatality mostly where physical facilities were not present (i.e., No Junction). Interestingly, except for the alignment at other junctions, the descriptor indicates opposite tendency results from the actual data. Concerning pedestrian physical facilities, like the road type, the results described tendency to a chance of fatality mostly happen where physical pedestrian facilities were not present. However, for 2018 data statistics of this nature were not found. This because, it was felt that statistics from 2018 regarding pedestrian road-crashes at pedestrian physical facilities conclude assist interpreting more road-crash tendencies more objectively. Although detailed data statistics from 2018 on pedestrian physical facilities were not found, with data statistics from 2017, the FCM prototype describes an indicative a decreasing tendency on the chance of fatality risk in the case of absent pedestrian physical facilities. These results are rather promising. With respect to the system sensitivity response, two additional scenario rules (i.e., FCM descriptor data split into 75% and 85% sample ratios) will be explored in Chapter 9 (Lanera et al., 2019; Saha et al., 2021). Further research work is this on exploring additional scenarios is highly suggested.

8.5. Observations

The findings from this chapter demonstrate several contributions that can improve pedestrian safety. This because, the FCM proves its potential to systematically describe relevant information from the investigated road elements. Furthermore, the model was calibrated with different sets of rules [see Section 8.2]. The FCM verification demonstrator provided a basis for investigating different regions: GB, and London. The data sample was validated using actual road data. The data was then compared with information from official sources (DfT, 2017. When compared with data from 2018, in overall for GB, the trends are in alignment with the data from the official sources (DfT, 2017. When looking into more detailed information the demonstrator to fails to precise with information on the tendency of fatalities by (ca. 10%). On the contrary for London, the trends are misaligned with the data from the official sources. This could be due to a difference in the size of the data sample used for each area (i.e., GB: 470, London: 131). The findings of the FCM verification demonstrator shows in overall, for GB, most accidents by road type, occurring and being projected to occur when physical facilities were/are not present. Chapter 5 agrees. Moreover, for London by local authority, the FCM verification demonstrator shows most accidents in Enfield (with 10 records) (DfT, 2017). Interestingly, the system demonstrated a decreasing tendency which aligned with the findings from actual statistical data from official sources from 2018. Therefore, the FCM prototype proved to be a basis for examining GB road data. Turning into the knowledge discovery, in overall, the results from the FCM prototype present some promising findings. More in detail, from the same analysis, the FCM prototype distinguishes three separable types of information: pedestrian general attributes (i.e., gender, age group); road environment (i.e., weather, speed limits, lighting); road infrastructure (i.e., road type, pedestrian physical facilities). Though the analysed data show lower accuracy rates in Section 8.4, still the FCM prototype proves its applicability [see Sections 8.2, and 8.3]. Therefore, to improve the knowledge discovery procedure, harmonising data quality through the incorporation of more sophisticated algorithms is highly recommended (DfT, 2017; DfT, 2018).

9. Sensitivity Testing

9.1. Introduction

This chapter tests the FCM prototype sensitivity response compared with actual data from 2018 (DfT, 2017; DfT, 2018) [see Chapter 8]. The main purpose of the chapter is to analyse a selection of parameters and test the level of accuracy of the FCM sensitivity response, when using different rules. The chapter is organised in five main sections. Section 9.1 shows the overall structure of the chapter. Section 9.2 defines new set of assumptions to compare statistic data with the trained data by the FCM prototype. Section 9.3 compares statistical data with the FCM prototype using new data rules. This to assess and quantify the sensitivity response. Section 9.4 validates the findings with expert opinion. Section 9.5 discusses the findings.

9.1.1. Overall Approach

Regarding the sensitivity response of predictive modelling, several assessment approaches have been previously developed. For example, Assi et al. (2020) agree that the sensitivity response of a model, can be calculated based on the ratio of road-crashes correctly predicted, and the total number of actual road-crashes. However, the authors also claim that despite the FCM implementation simplicity, the initialisation of a prototype remains a key issue. For instance, compared with the other developed predictive models the SVM-FCM model has outperformed its accuracy. Therefore, the association of this model with more sophisticated models being previously proposed (Mocan and Dioşan, 2016; Assi et al., 2020). This chapter entailed a systematic investigation of the FCM sensitivity response to actual data when using different rules (DfT, 2017; DfT, 2018). The different rules used are the model response to 85% and 75% compared to 80% [see Section 3.3.1]. To assess the FCM sensitivity response, primarily a given set of assumptions (i.e., set criteria) (Marisamynathan and Vedagiri, 2019) was specified. These assumptions were then tested on a given distinct number of pedestrian fatality rates. Additionally, new rules were applied to different variables: dependent (i.e., pedestrian fatalities), and independent (i.e., pedestrian general attributes, road environment, road infrastructure).

9.2. Criteria Specification

The criteria specification to examine the FCM sensitivity response, accounted for assumptions on the number of accidents rates, and new rules applied to investigate actual road data from 2017. The set criteria considered the following three assumptions:

i. Assumption 1 - Number of pedestrian road-crash casualties: 500.

ii. Assumption 2 - Number of pedestrian road-crash casualties: 250.

iii. Assumption 3 - Number of pedestrian road-crash casualties: 100.

The above assumptions were proposed as changes for the input variables, while the values of other variables were kept unchanged. From this exercise, the results might likely be uncertain. The range of input values of the investigated variables was based on the minimum and maximum values determined from actual data. The unchanged values of other variables were calculated as the estimated means of the data set under examination (DfT, 2017; DfT, 2018). The output parameters, i.e., dependent variables are given by [see Chapter 8]:

a. Gender.

b. Age group.

c. Weather.

d. Speed limits.

e. Lighting.

f. Road type.

g. Pedestrian physical facilities type.

FCM Demonstrator

With respect to the FCM demonstrator as per Section 8.2, in addition to the statistics, both the descriptor and the predictor were explored. Additionally, it included verifying two additional partition scenarios as follows:

Scenario 1: statistics 100%, descriptor 50%, predictor 50%.

FCM Prototype

For the FCM demonstrator in Section 8.2 where the statistics where contrasted with both the descriptor and the predictor were examined. Since the descriptor scored a higher accuracy than the predictor, for the FCM prototype the statistics where only compared with the descriptor [see Section 8.3]. The knowledge discovery it included verifying the following modes:

Scenario 2: statistics 100%, descriptor 80%.

FCM Sensitivity Response

Similar to the FCM prototype in Section 8.3, for the FCM sensitivity response the following scenario rules for the descriptor were tested:

Scenario 3: statistics 100%, descriptor 85%.

Scenario 4: statistics 100%, descriptor 75%.

9.3. Co-relational Analysis: Short-Term Projections

The co-relational analysis assessed the FCM sensitivity response to actual road data from 2017 (i.e., 470 fatalities) using two new sets of rules: *Scenario 3, and Scenario 4* [see Section 9.2]. The FCM sensitivity response validation procedure shows for *Scenario 3*, the new rules on 85% describing 400 fatalities. On *Scenario 4*, the new rules on 75% describing 353 fatalities (Lanera et al., 2019; Saha et al., 2021). However, it is relevant to mention that, compared with the overall statistical analysis in Section 6.4, the standalone data set represented in Table 6.1 differs by 40 records. This is a key issue as it might quite significantly affect the interpretation of examined the data. Similar to the knowledge discovery carried out in Section 8.3, the sensitivity response verification was carried out on the following factors: gender, age group, weather, speed limits, lighting, road type and pedestrian physical facilities.

9.3.1. Pedestrian General Attributes: Sensitivity Response by Gender

Figure 9.1 contrasts by gender the statistics from 2017, the FCM prototype (i.e., descriptor 80%), and the FCM sensitivity response (i.e., descriptor 85%) and the statistics from 2018.

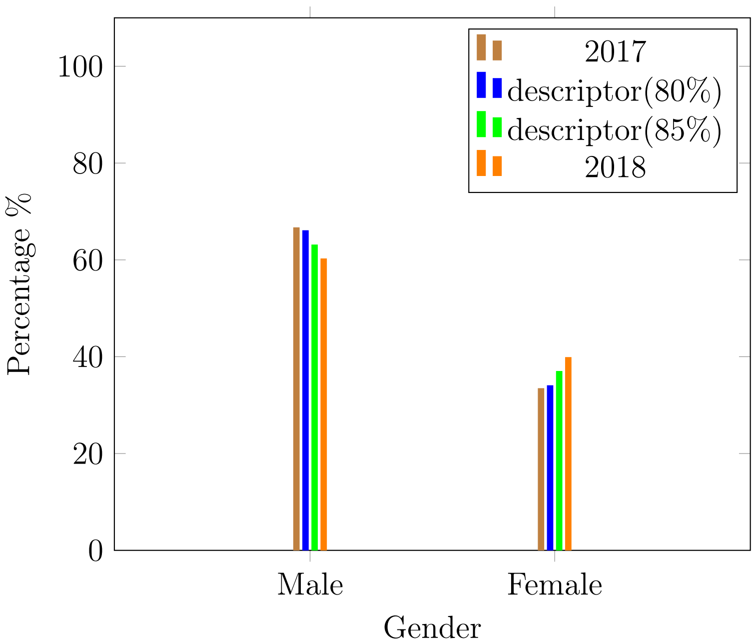


Figure 9.1.: FCM sensitivity response by gender: statistics (2017, 2018) vs. descriptor (80%, 85%)

The results in Figure 9.2 show that the sensitivity response (85%) and the statistics from 2018 differ by around 2.84%. As result, compared with the FCM prototype (80%) its accuracy increases [see Section 8.3].

Figure 9.2 contrasts by gender the statistics from 2017, the FCM prototype (i.e., descriptor 80%), the FCM sensitivity response (i.e., descriptor 75%) and the statistics from 2018.

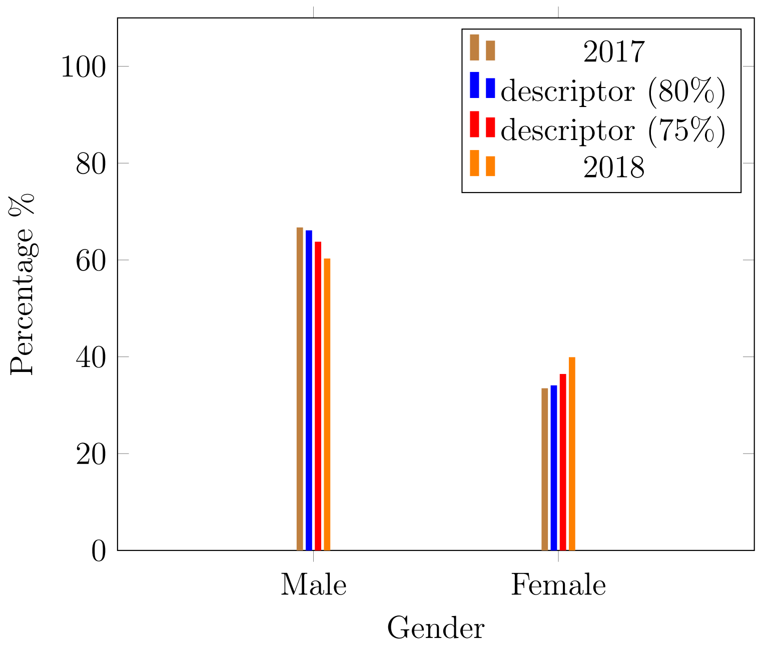


Figure 9.2.: Sensitivity response by gender: statistics (2017, 2018) vs. descriptor (80%, 75%)

The results between Figure 9.2 the FCM sensitivity response (75%) and the statistics from 2018 differ by around 3.37%. As result, compared with the FCM prototype (80%) its accuracy increases [see Section 8.3].

9.3.2. Pedestrian General Attributes: Sensitivity Response by Age Group

Figure 9.3 contrasts the statistics from 2017, the FCM prototype (80%), the FCM sensitivity response (85%) and the statistics from 2018, by age group[1].

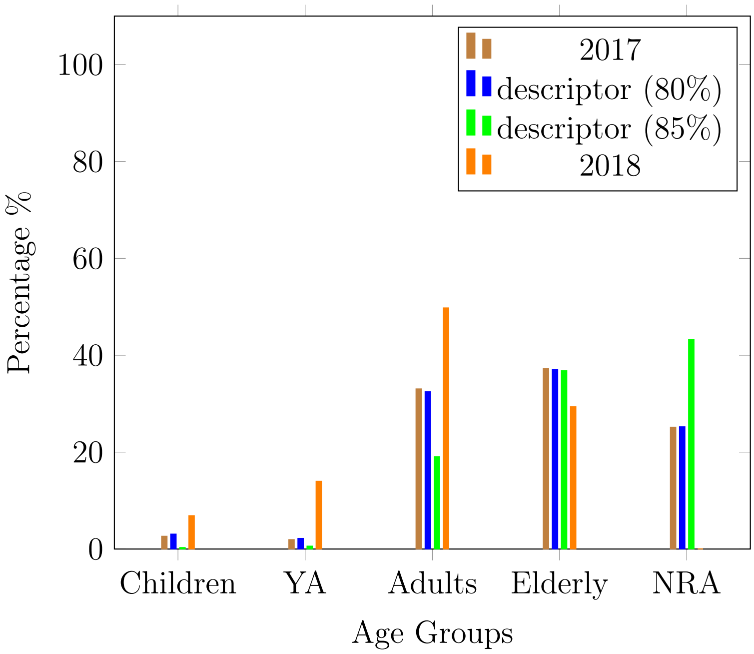


Figure 9.3.: Sensitivity response by age group: statistics (2017, 2018) vs. descriptor (80%, 85%)

Figure 9.3 shows that except for the elderly, the FCM sensitivity response (i.e., descriptor 85%) is considerably less accurate than the prototype (i.e., descriptor 80%) [see Section 8.3].

[1] Children (0-14); Young adolescents (YA) (15-24); Adults (25-64), Elderly (65+); Non recorded ages (NRA)

Figure 9.4 contrasts the statistics from 2017, the FCM prototype (80%), the FCM sensitivity response (75%) and the statistics from 2018, by age group[2].

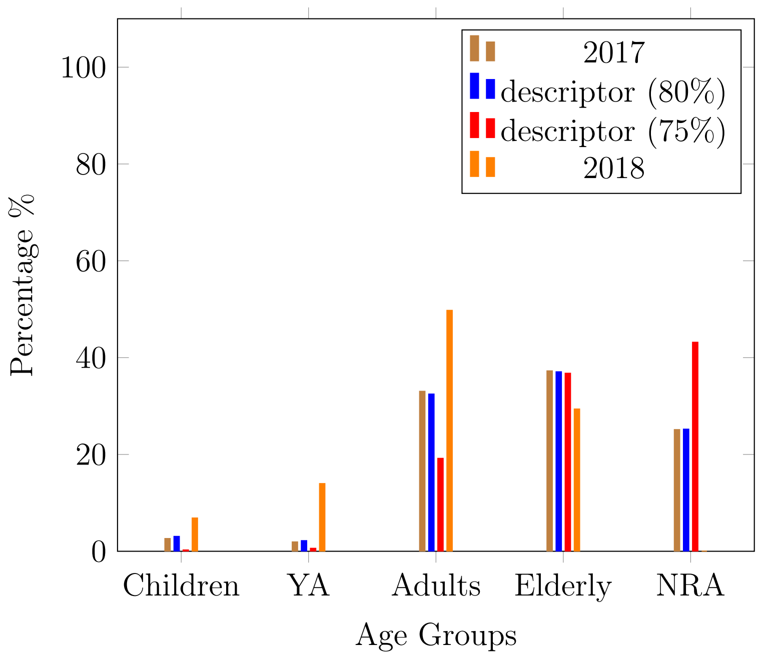


Figure 9.4.: Sensitivity response by age group: statistics (2017, 2018) vs. descriptor (80%, 75%)

Similar to Figure 9.3, Figure 9.4 shows that except for the elderly, the sensitivity response (75%) is considerably less accurate than the prototype (80%) [see Section 8.3].

[2]Children (0-14); Young adolescents (YA) (15-24); Adults (25-64), Elderly (65+); Non recorded ages (NRA)

9.3.3. Road Environment: Sensitivity Response by Weather

Figure 9.5 contrasts the statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (85%) and the statistics from 2018, by weather[3].

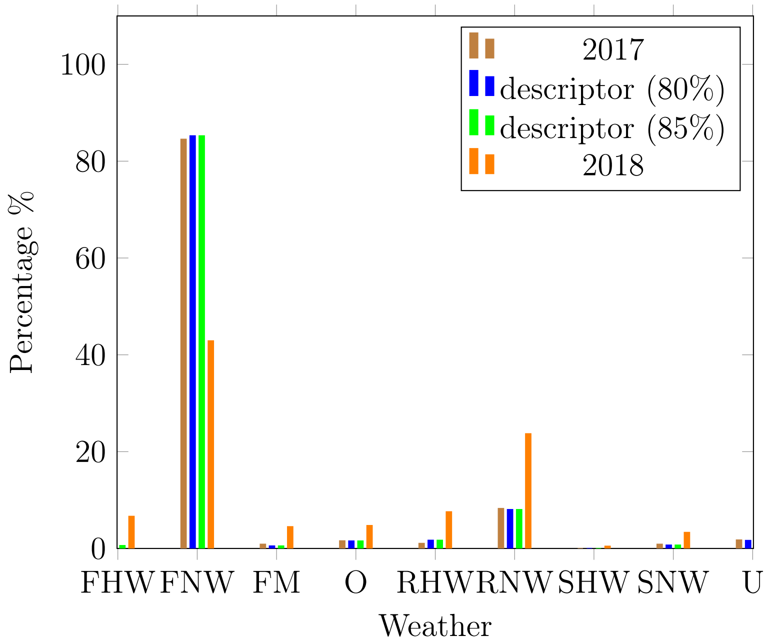


Figure 9.5.: Sensitivity response by weather: statistics (2017, 2018) vs. descriptor (80%, 85%)

Similar to Figure 8.8, compared with the prototype, the FCM sensitivity response shows less accurate results than the actual road data (DfT, 2017; DfT, 2018) [see Section 8.3].

[3]Fine-high winds (FHW); Fine-no winds (FNW); Fog or mist (FM); Other(O); Raining-high winds (RHW); Raining-no winds (RNW); Snowing-high winds (SHW);Snowing-no winds (SNW); Unknown (U)

Figure 9.6 contrasts statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (i.e., descriptor 85%) and the statistics from 2018, by weather[4].

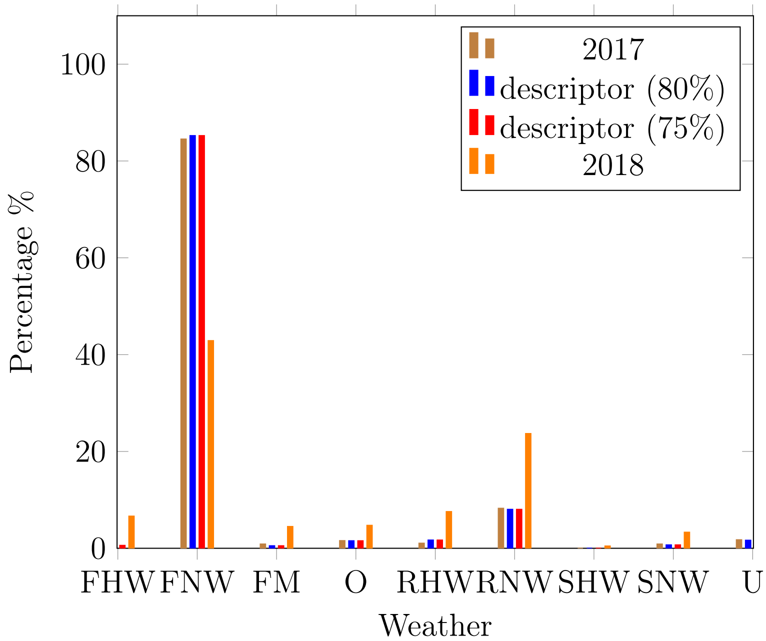


Figure 9.6.: Sensitivity response by weather: statistics (2017, 2018) vs. descriptor (80%, 75%)

Similar to Figure 9.5, Figure 9.5 shows that compared with the prototype the FCM sensitivity response with lower accuracy results [see Section 8.3].

[4]Fine-high winds (FHW); Fine-no winds(FNW); Fog or mist (FM); Other(O); Raining-high winds(RHW); Raining-no winds(RNW); Snowing-high winds (SHW); Snowing-no winds (SNW); Unknown (U)

Road Environment: Sensitivity Response by Speed Limits

Figure 9.7 contrasts the statistics from 2017, the FCM prototype (80%), the FCM sensitivity response (85%) and the statistics from 2018, by speed limits [5].

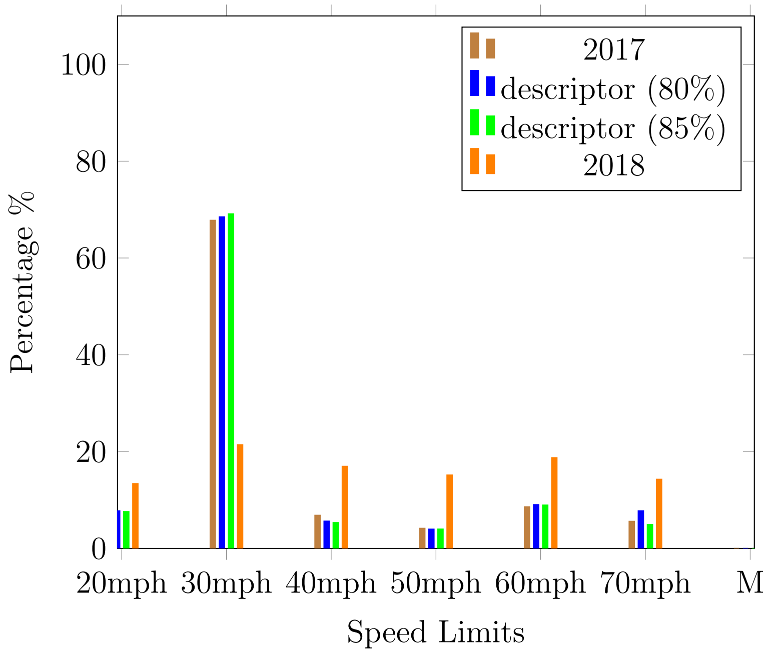


Figure 9.7.: Sensitivity response by speed limits: statistics (2017, 2018) vs. descriptor (80%, 85%)

Compared with the prototype in Figure 8.7, Figure 9.7 shows the FCM sensitivity response with lower accuracy results in relation the road actual data, especially for speed limits set at 70 mph (i.e., 2.84%) [see Section 8.3].

[5] 1-20 mph, 21-30 mph, 31-40 mph, 41-50 mph, 51-60 mph, 61-70 mph, Motorway(M)

Figure 9.8 contrasts statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (75%) by speed limits[6].

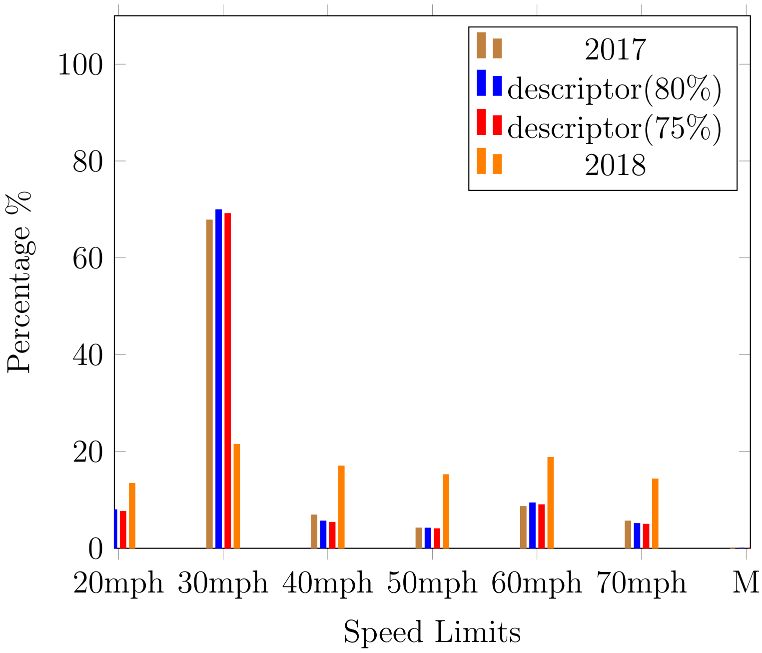


Figure 9.8.: Sensitivity response by speed limits: statistics (2017, 2018) vs. descriptor (80%, 75%)

Compared with the prototype in Figure 8.7, Figure 9.8 presents the FCM sensitivity response, with lower accuracy results in relation to the actual road data. For speed limits set at 20 mph a slight improvement was noticed (0.79%) [see Section 8.3].

[6] 1-20 mph, 21-30 mph, 31-40 mph, 41-50 mph, 51-60 mph, 61-70 mph, Motorway(M)

9.3.4. Road Environment: Sensitivity Response by Lighting

Figure 9.9 contrasts the statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (85%) and the the statistics from 2018, by lighting[7].

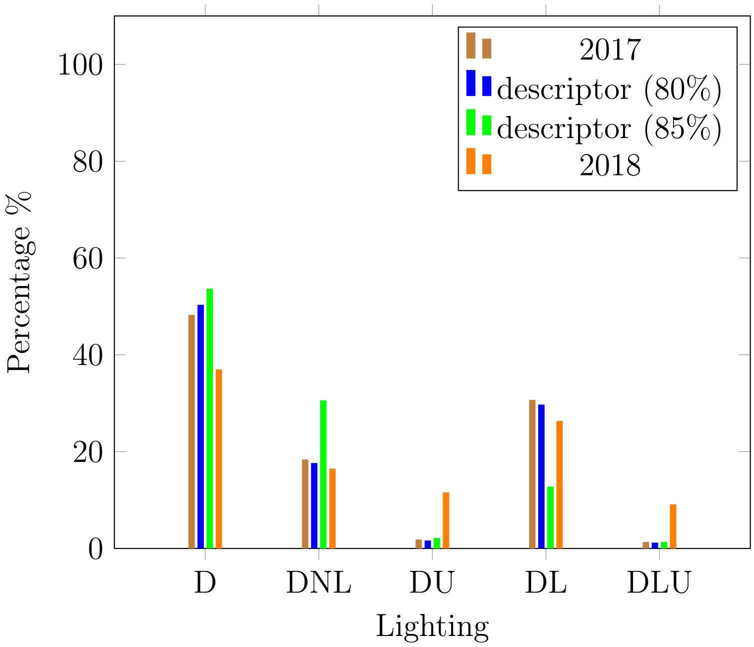


Figure 9.9.: Sensitivity response by lighting: statistics (2017, 2018) vs. descriptor (80%, 85%)

Compared with the prototype in Figure 8.8, Figure 9.9 presents the FCM sensitivity response with lower accuracy results. Interestingly, at night (i.e., DL) a slight improvement was noticed (0.49%) [see Section 8.3].

[7]Daylight (D); Dark no lighting (DNL); Dark unlit (DU); Dark lit (DL); Dark light unknown (DLU)

Figure 9.10 contrasts the statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (75%) and the statistics from 2018, by lighting[8].

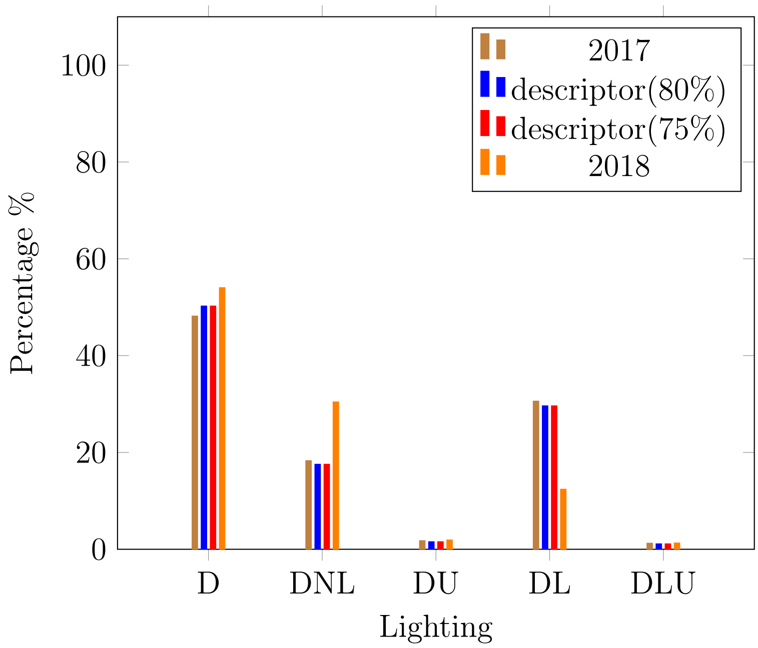


Figure 9.10.: Sensitivity response by lighting: statistics (2017, 2018) vs. descriptor (80%, 75%)

Compared with the prototype, Figure 8.8, in Figure 9.10 the FCM sensitivity response shows similar accuracy results than the actual road data [see Section 8.3].

[8]Daylight (D); Dark no lighting (DNL); Dark unlit (DU); Dark lit (DL); Dark light unknown (DLU)

9.3.5. Road Infrastructure: Sensitivity Response by Road Type

Figure 9.11 contrasts the statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (85%) and the statistics from 2018, by road type[9] .

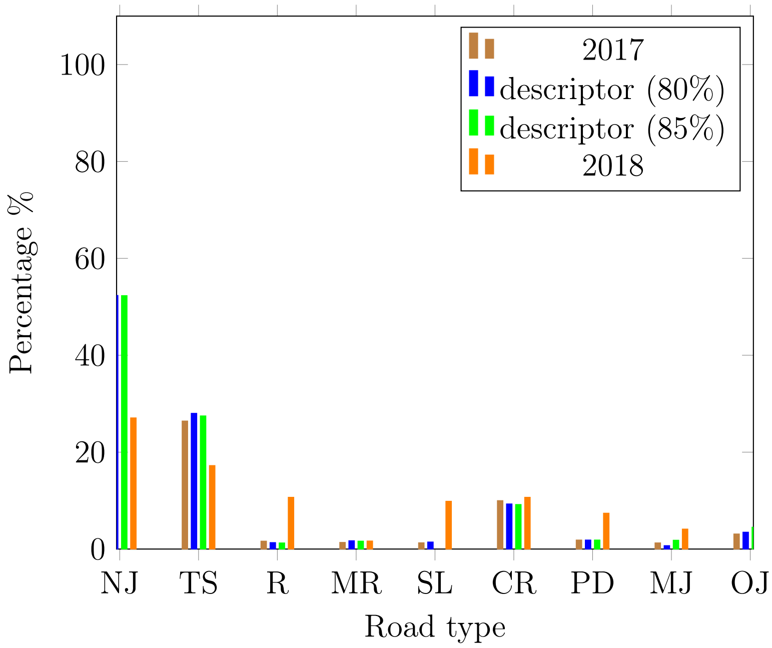


Figure 9.11.: Sensitivity response by road type: statistics (2017, 2018) vs. descriptor (80%, 85%)

Compared with the prototype in Figure 8.9, Figure 9.11 shows the FCM sensitivity response with similar accuracy results in relation to the actual data (STATS19; DfT, 2019). Interestingly, additionally it presents slight more accurate results at staggered junctions (0.07%) and at multiple junctions (1.12%) [see Section 8.3].

[9] 0- No Junction (NJ); 1- T staggered (TS); 2- Roundabouts (R); 3- Mini-roundabouts (MR); 4- Slip road (SL); 5-Cross roads (CR); 6-Private drive(PD) 7-Multiple Junction (MJ); 8-Other Junction (OJ)

Figure 9.12 contrasts the statistics from 2017, the FCM prototype (80%), and the FCM sensitivity response (75%) and the statistics from 2018, by road type[10].

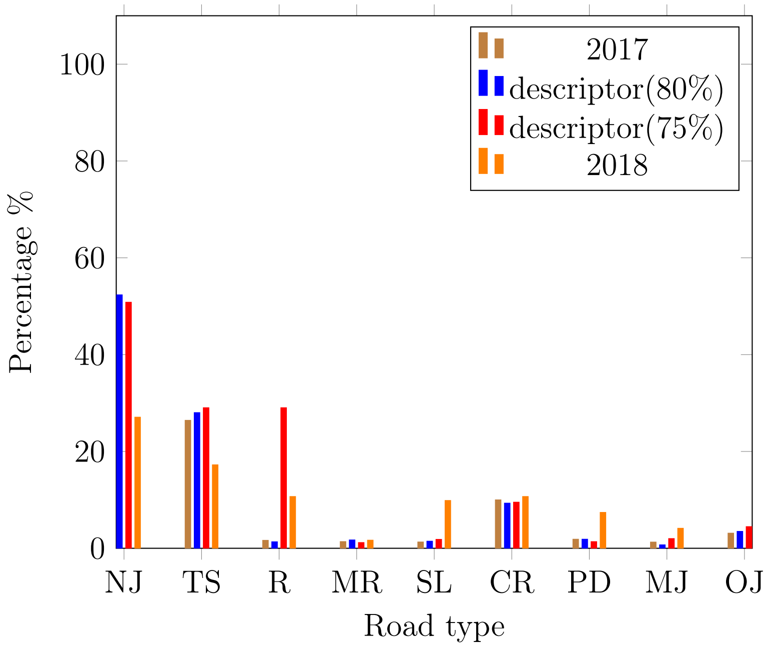


Figure 9.12.: Sensitivity response by road type: statistics (2017, 2018) vs. descriptor (80%, 75%)

Compared with the FCM prototype, in Figure 9.12 the FCM sensitivity response shows similar accuracy results. Interestingly, in addition to that it presents significantly accurate trends at roundabouts and slightly less accurate at crossroads (0.2%), and also at multiple junctions (1.12%) [see Section 8.3].

[10] 0- No Junction (NJ); 1- T staggered (TS); 2- Roundabouts (R); 3- Mini-roundabouts (MR); 4- Slip road (SL); 5-Cross roads (CR); 6-Private drive (PD) 7-Multiple Junction (MJ); 8- Other Junction (OJ)

9.3.6. Road Infrastructure: Sensitivity Response by Pedestrian Facilities

Since actual data from 2018 on pedestrian physical facilities[11] was not found, Figure 9.13 contrasts statistics from 2017, the FCM prototype (i.e., descriptor 80%) and the FCM sensitivity response for two distinct rules: 85%, 75%.

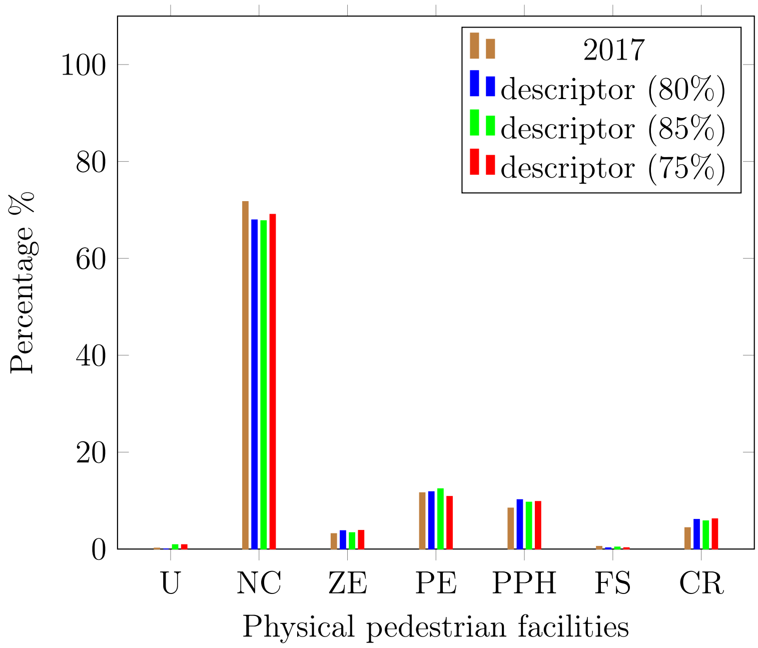


Figure 9.13.: Sensitivity response by pedestrian physical facilities: statistics (2017) vs. descriptor (80%, 85%, 75%)

Compared with the FCM prototype, the FCM sensitivity response indicates a decreasing tendency at the FCM descriptor 85% zebra, pelican, pedestrian phase, and central refuge. On the other hand, the FCM sensitivity response indicates a decreasing tendency at the FCM descriptor 75% at zebra, pedestrian phase, and the central refuge [see Section 8.3].

[11] Unknown(U); None crossing within 50 m(NC); Zebra (ZE); Pelican (PE); Pedestrian Phase (PPH); Footbridge or subway (FS); Central refuge (CR)

9.3.7. Lessons Learnt: Sensitivity Analysis

The FCM sensitivity response was tested on two different scenario rules: 85%, and 75%. The results were compared with the FCM prototype (80%) and actual statistics from 2018 (i.e., descriptor 80%). By gender, like the FCM prototype, the FCM sensitivity response succeeds with the indicative the trends. Moreover, the 75% set of rules represents the highest accuracy rate. Regarding the age group, the FCM sensitivity response partially succeeds with the indicative trends. For the elderly, the 75% set of rules represents the highest accuracy rate. One reason for misalignment could be due lacking information from attributes such as 'non-recorded age' leading to mislead the system. As regards weather, the FCM sensitivity response mostly fails with the indicative trends. Future work on using of more sophisticated algorithms is therefore suggested. With respect to speed limits, the FCM sensitivity response mostly fails with the indicative trends. Interestingly, for speed limits set at 70 mph, the 85% set of rules represents the highest accuracy rate. On lighting, the FCM sensitivity response mostly fails with the indicative trends. Alike on weather, future work on using more sophisticated algorithms is proposed. Concerning the road type, the FCM sensitivity response mostly fails with the indicative trends. Nevertheless, for roundabouts, the 75% set of rules represents promising trends. In terms of pedestrian physical facilities, disappointingly data was found from neither data from 2018 nor from 2019. As the most relevant factor in this study, a further analysis on how to access this type of data is suggested. This is because the data from 2017 and the scenario rules distinguish accidents to occur in the zebra most likely (slightly on the 75%), the pelican (slightly more on the 85%), pedestrian phase (slightly more on the 85%) and central refuge (slightly more on the 75%).

9.4. Verification of the conducted Research: Expert Opinion

This study mostly focused on adopting engineering measures to assess pedestrian safety [see Chapters 2, 4 and 5]. To enable a deeper understanding of the study under research and advance the collective insights, qualitative research methods were applied by adding complementary contributions from a designated group of experts to reduce bias (Fischeret al., 2023). Therefore, in this section it is aimed to demonstrate the power and role of qualitative research methods and how these can be employed. Here, two types of qualitative research techniques were applied: focus groups and questionnaires (Savin-Banden et al., 2023). These were used to validate the findings with the opinion of a group of experts (i.e., engineers) (Mirhashemi et al., 2022). The questions were deployed to investigate nonnumerical methods, as to learn nonquantifiable phenomena from the respondents (Savin-Banden et al., 2023). The knowledge from the experts was collected in the form of an online questionnaire (Hakkanen et al., 2022; Muchanga-Hvelplund, 2023). Furthermore, the deployed questionnaire was created with a view of formulating a novel set of exploratory questions (Anderson et al., 2022).

9.4.1. Background of the Experts

The knowledge of the experts (e.g., road designers and engineers) can provide with robust and reliable information. Thus, a fundamental contribution for this research study. In addition to that, it also adds human value to the understanding in which people in different parts of the world with a similar professional background and experience, perceive pedestrian safety. The focus group questionnaire was conducted on 19 participants. The participants were located in 16 countries around the world

(Australia: 5.3%; Croatia: 10.5%; Denmark: 10.5%; Ecuador: 5.3%; Ghana: 5.3%; Malaysia: 10.5%; Moldova: 5.3%; Nigeria: 5.3%; North Macedonia: 5.3%; Portugal: 5.3%; Romania: 5.3%; Serbia: 5.3%; Spain: 5.3%; Tunisia: 5.3%; Turkey: 5.3%; United Kingdom: 5.3%). The participants were affiliated to distinct institutions (i.e., Academia: 42.1%; Public Sector: 36.8%; Private Sector: 42.1%; Other: 5.3%). Among others, some of these institutions included:

i. Academia - Faculty of Urbanism and Architecture; Monash University Malaysia University of Zagreb.

ii. Public Sector - Malaysian Institute of Road Safety Centre; United Nations office for Project Services (UNOPS).

iii. Private Sector - The World Bank.

9.4.2. Questionnaire

This questionnaire investigates the applicability of AI to assessing pedestrian safety in Great Britain (GB). The aim is to validate the findings of the deployed prototype with the expert's opinion by comparing findings from questionnaire with the results of the AI data analysis indicative trends. The questionnaire study was conducted with 19 experts. Confirmatory road risk factor analysis differentiated pedestrian accidents leading to fatalities into four main categories: human, environmental, infrastructure, and data quality. The questionnaire includes in total 21 questions that are split into three main parts. Table 9.1 illustrates the survey structure (Muchanga-Hvelplund, 2023).

Table 9.1.: Questions

Part I
1. Please indicate your affiliation
2. Please select your country of residence
Part II - Examined road elements leading to pedestrian road-crash casualties
1. To your knowledge, rank the importance of human, environmental, and infrastructure factors that affect pedestrian road-crash casualties, where 1 means irrelevant and 5 means critical. Human, Environmental, Infrastructural, The quality of the data)
2.Can you write down up to 10 words that come to mind when you think about pedestrian safety.
Part III - Validation of the findings: The projected data on GB road-crash fatalities for 2018 using 2017 data are presented below, by category (i.e., Human, Environment, Infrastructure).
1. *Human category:* How confident are you that the anticipated GB percentages for 2018 match the actual data in your region for 2018?
2. *Road environment category:* How confident are you that the anticipated GB percentages for 2018 match the actual data in your region for 2018?
3. *Road infrastructure category:* How confident are you that the anticipated GB percentages for 2018 match the actual data in your region for 2018?
4. To reduce pedestrian road-crash casualties, based on the analysis of GB road data from 2017 and your region's data, which factor would you prioritise evaluating?
5. In terms of pedestrian general attributes (i.e., gender, age group) the research findings indicate human error as a major factor in pedestrian road-crash fatalities. Can you explain ways you have for addressing both human error factors affecting pedestrian safety? Give some examples of concrete effective measures (e.g., by gender and age group).
6. In terms of the road environment, the findings indicate that adverse environmental conditions (e.g., weather), speed limits, and lighting systems are major influencing factors in pedestrian road-crash fatalities. How confident are you that these environmental conditions affect pedestrian safety?
7. Can you provide with recommendations with concrete examples, for reducing the impact of road environmental conditions such as rain and wind, permanent speed limits set at 30 mph, and lower levels of lighting that result in pedestrian road-crash casualties?
8. Road infrastructure results show that the absence of Pedestrian physical facilities (e.g., footbridge, zebra crossings) is a critical factor affecting pedestrian road-crash fatalities. How confident are you that absence pedestrian physical facilities affect pedestrian safety?
9. Can you provide with concrete measures that can mitigate the impact of absent pedestrian physical facilities leading to pedestrian road-crash fatalities?
Part IV - Validation of the recommendations based on the findings from the examined road data.
1. The findings of this show a paucity of literature on various levels of cognitive disabilities affecting pedestrian road crashes. What would you suggest to prepare for future research on this topic?
2. To date, fewer studies have focused on gathering road data from contributory elements that can help gain new insights to reduce road-crash fatalities. Can you provide with concrete recommendations to address the data collection procedure?
3. Overall, how confident are you that the above questions are applicable to other regions (e.g., regional, international)?

9.4.3. Findings from the Experts

In this section the opinion from experts will validate the findings from the deployed prototype. Figure 9.14 shows the views of the participants through ranking the importance of human factors affecting pedestrian road-crash casualties, where 1 means irrelevant and 5 means critical.

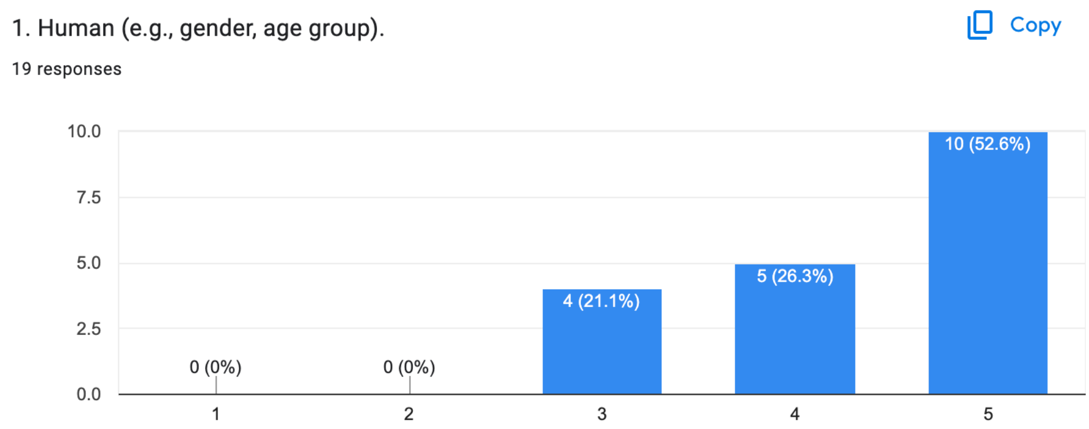


Figure 9.14.: Ranking of human factors affecting pedestrian safety

As presented in Figure 9.14, most participants indicated the importance of human factors that affect pedestrian road-crash casualties to be five (52.6%) followed by four (26.3%) and lastly three (21.1%). By contrast, when considering some human attributes (i.e., gender, age groups) the prototype demonstrated to align the actual data (i.e., 2018) ahead in time by gender. Moreover, it detailed a rank on the critical road-crash fatalities patterns accordingly: male and elderly. However, by age group, in terms of the prototype accuracy it does not present acceptable performance measures (i.e., MOE) [see Section 8.3].

Figure 9.15 below presents the participants ranking on the importance of environmental factors affecting pedestrian road-crash casualties, where 1 means irrelevant and 5 means critical.

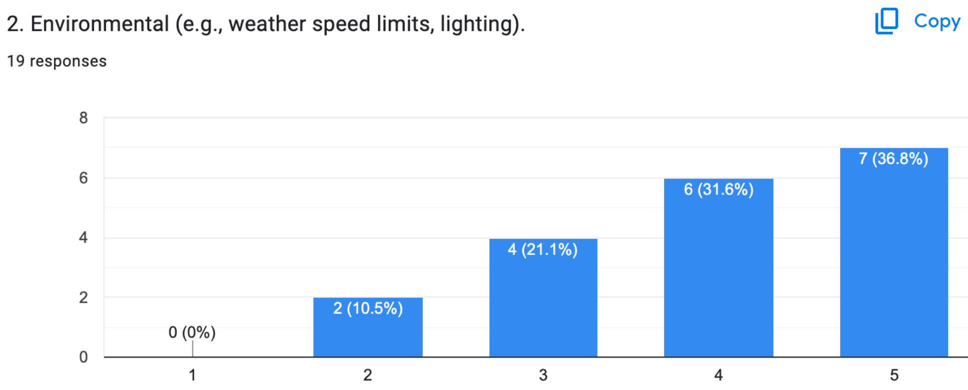


Figure 9.15.: Environmental factors affecting pedestrian safety

As shown in Figure 9.15, participants in the group indicated that the importance of environmental factors that affect pedestrian road-crash casualties is as follows: five (36.8%), four (31.6%), three (21.1%), and two (10.5%). In contrast, when considering some environmental attributes (i.e., speed limits, weather, lighting condition) the prototype ranked in detail road-crash casualty patterns as follows: fine weather with no winds, permanent speed limits set at 30 mph during, with daylight [see Section 8.3]. However, on the weather the prototype indicates opposite results to the indicative trends. Regarding the speed limits, the prototype does not present an acceptable accuracy performance measure (i.e., MOE). In addition to that, as regards the lighting condition, the prototype does not present acceptable accuracy performance measures (i.e., MOE, z-score) [see Section 8.3].

Figure 9.16 shows the participants ranking on the importance of infrastructural factors affecting pedestrian road-crash casualties, where 1 means irrelevant and 5 means critical.

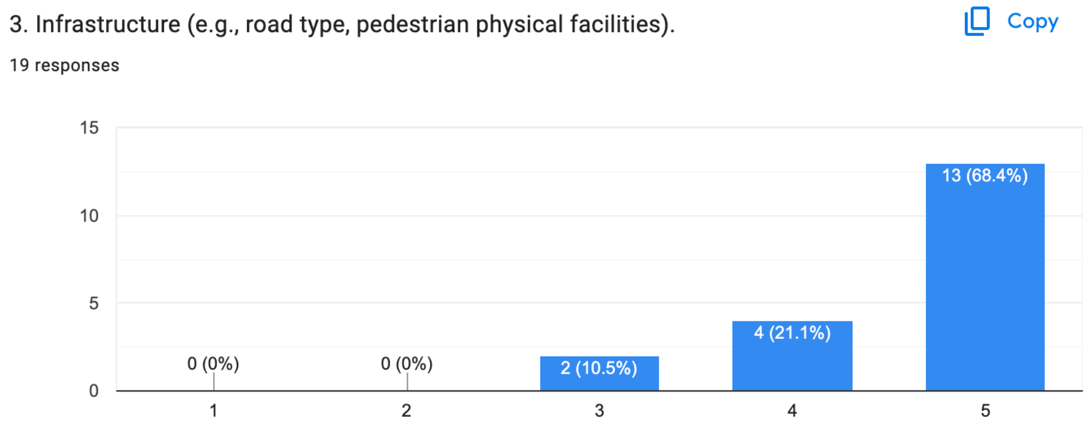


Figure 9.16.: Infrastructural factors affecting pedestrian safety

As can be seen in Figure 9.16, most of the participants indicated the importance of infrastructural factors affecting pedestrian road-crash casualties to be five (68.4%), followed by four (21.1%) and three (10.5%). On the other hand, based on a few physical infrastructural attributes (i.e., road type, pedestrian physical facilities), the prototype prioritised road-crash casualty patterns at no pedestrian physical facility within 50 m, followed by at no junction. However, about the pedestrian physical facilities, actual statistics from 2018 were not found. In addition to that, the road type prototype mostly indicates opposite results to the indicative trends with the critical addition that the prototype does not present acceptable accuracy performance measures (i.e., MOE) [see Section 8.3].

Figure 9.16 introduces the ranking of the participants on the importance of the quality of the data affecting pedestrian road-crash casualties, where 1 means irrelevant and 5 means critical.

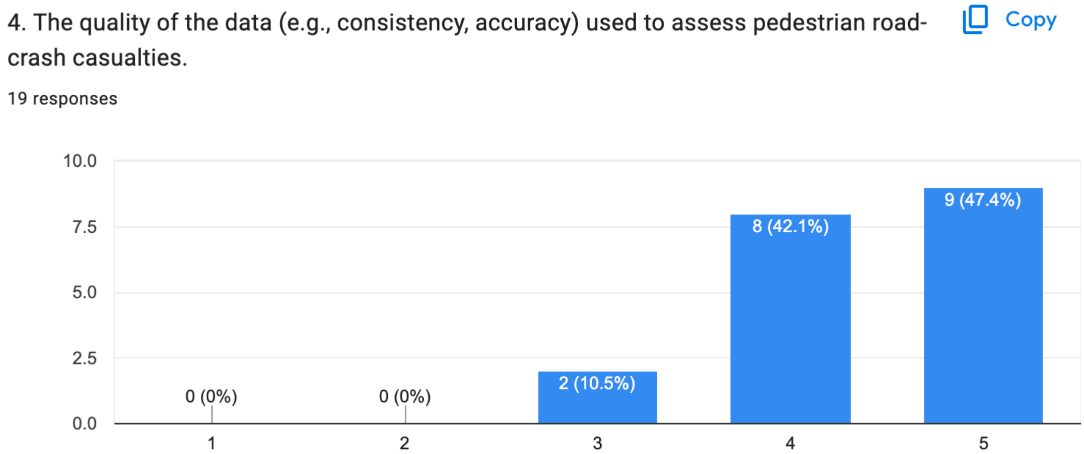


Figure 9.17.: Data quality affecting pedestrian safety

In Figure 9.17, most responses of the ranking regarding the importance of the quality of the data affecting pedestrian road-crash casualties indicated five (47.4%) followed by four (42.1%) and lastly three (10.5%). As presented above in Figures 9.15, 9.16, 9.17, based on a distinct of attributes (i.e., human, environmental, infrastructural), the prototype forms different road-crashes patterns [see Section 8.3]. One critical aspect previously mentioned, is the high rate of non-recorded ages, as it may likely affect the metrics of performance. Another critical element likely leading to system validation is disqualification at pedestrian physical facilities, is the lack of data availability in 2018 for road-crash fatalities in that specific location [see Section 8.3.3].

Figures 9.18 illustrates an overview of words related to pedestrian safety written by the participants.

5. Can you **write down** up to 10 words that come to mind when you think about pedestrian safety.

14 responses

VRU, visibilitiy, risk behaviour, mobile phone use,

Training, retraining, advocate visit to schools,zebra crossing, traffic light, pedestrian road Mark, traffic islands, foot bridges, subways, changing the attitudes of drivers

Sidewalk, illumination, pedestrian crossing, low traffic speed

Life, cost, economic, sidewalk, urban design, handrail, slopes, material

Pedestrian safety is paramount in promoting sustainable transport and assisting governments in achieving their net-zero targets. In addition to that, as an aging community is growing, in the last decade, more and more older pedestrians have been involved in pedestrian crashes. Hence pedestrian safety programs need to be customised for that specific age groups

car parked on sidewalks; kerbsides to along the ground; physical separation for the pedestrians.

Visibility, errors, pedestrian infrastructure, speed

Figure 9.18.: Pedestrian safety keywords

The words presented in Figure 9.18, illustrate the general perception of pedestrian safety from a group of experts. The presented words appeared to construct a meaning based on words 'visibility', 'age groups', 'speed' and 'pedestrian facilities'. In this question, neither the words 'data' nor standard were mentioned by the experts. By contrast, the prototype carried out a data-driven investigation on most of the keywords mentioned by the participants [see Chapter 2, Section 8.3].

Figure 9.19 presents the opinion of the participants regarding the anticipated percentages of GB for 2018 (by gender and age group) that match the actual data in their respective region for 2018.

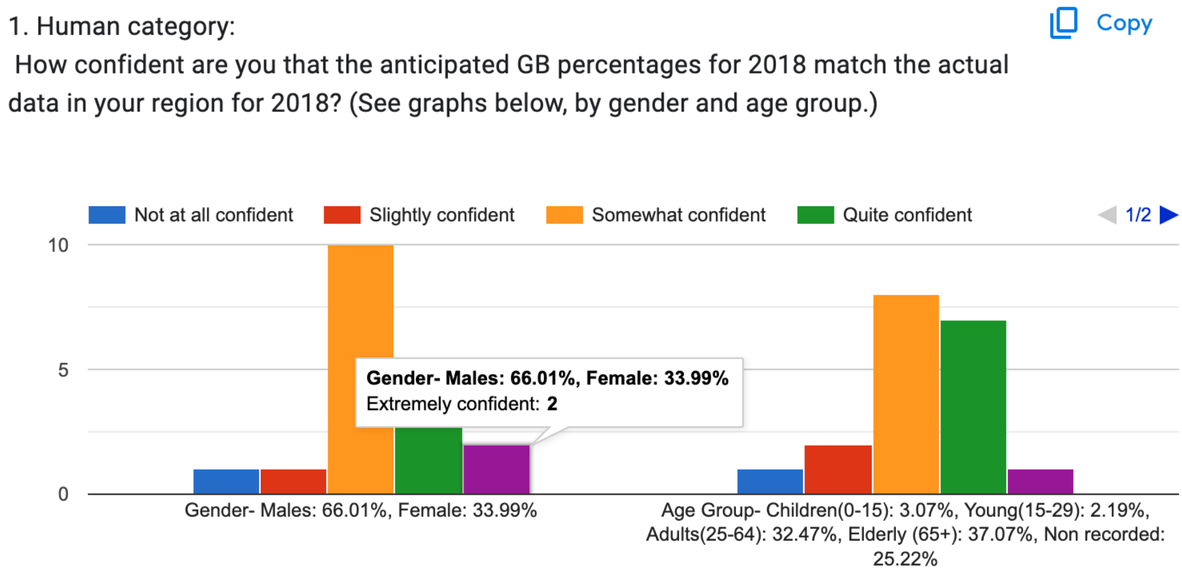


Figure 9.19.: Road-crash by gender and by age group: comparison between GB data and the participants' respective country of residence

Figure 9.19 shows many of the participants to confident with the projected trends by gender and age group. By gender, the results indicated somewhat confident (10 out of 19), followed by quite confident (5 out of 19), extremely confident (2 out of 19), slightly confident (1 out of 19), not at all confident (1 out of 19). By contrast, the prototype demonstrated to align the trends, by using data from 2017 [see Section 8.3]. The respondents indicated to have less confidence the anticipated percentages by age group where the results indicated somewhat confident (8 out of 19), followed by quite confident (7 out of 19), extremely confident (1 out of 19), slightly confident (2 out of 19), not at all confident (1 out of 19). In contrast, the results from the prototype demonstrated to align with the findings due to a significant number of non-recorded ages (NRA) [see Section 8.3.1].

Figure 9.20 presents the opinion of the participants on the anticipated GB percentages for 2018 (by weather, speed limits, and lighting systems) that match the actual data in their respective region, also for 2018.

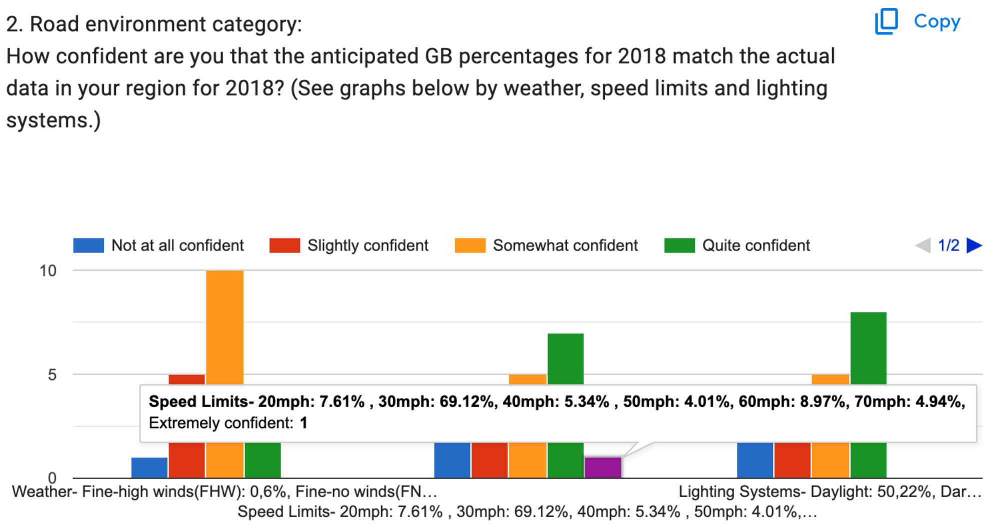


Figure 9.20.: Road-crash by weather, by speed limits, and by lighting systems: comparison between GB data and the respondents' respective country of residence

Figure 9.20 shows the responses of the participants on the level of confidence from the prototype projected trends. Regarding the speed limits results, these indicated quite confident (7 out of 18), followed by somewhat confident (5 out of 18), not at all confident (3 out of 18), slightly confident (2 out of 18), extremely confident (1 out of 18). On the other hand, the prototype the MOE does not present acceptable accuracy performance for measures. As regards lighting systems, the results indicated quite confident (8 out of 18), followed by somewhat confident (5 out of 18), slightly confident (3 out of 18), not at all confident (2 out of 18). By contrast, the prototype does not present acceptable accuracy performance measures (i.e., MOE, z-score). On the weather, the results indicated somewhat confident (10 out of 18), followed by slightly confident (5 out of 19), quite confident (3 out of 18), not at all confident (1 out of 19). In contrast, the prototype indicates opposite results to the indicative trends [see Section 8.3].

Figure 9.21 presents the opinion of the participants on the anticipated percentages of GB for 2018 (by road type and pedestrian facilities type) that corresponds to the actual data in their respective region also for 2018.

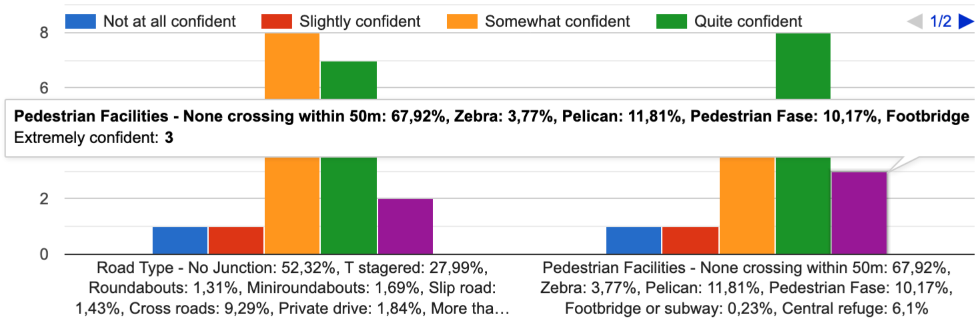


Figure 9.21.: Road-crash by road type and by pedestrian facilities type: comparison between GB data and the respondents' respective country of residence

Figure 9.21 presents the responses of the participants on the level of confidence from the prototype projected trends. Regarding the road type, these indicated somewhat confident (8 out of 19), followed by quite confident (7 out of 19), extremely confident (2 out of 19), slightly confident (1 out of 19), not at all confident (1 out of 19). In overall, the prototype indicates opposite results to the indicative trends and in addition to that, the prototype MOE does not present an acceptable performance for measuring accuracy. As regards pedestrian physical facilities, the results indicated quite confident (8 out of 19), followed by somewhat confident (4 out of 19), extremely confident (3 out of 19), slightly confident (1 out of 19), not at all confident (1 out of 19). Here, GB data from 2018 on road-crash fatalities at pedestrian physical facilities for comparison with the findings of the prototype was not found [see Section 8.3].

Figure 9.22 presents the priority ranking of the respondents on pedestrian road crash factors (i.e., human, environmental, infrastructure, other) of GB road data from 2017 and their respective region.

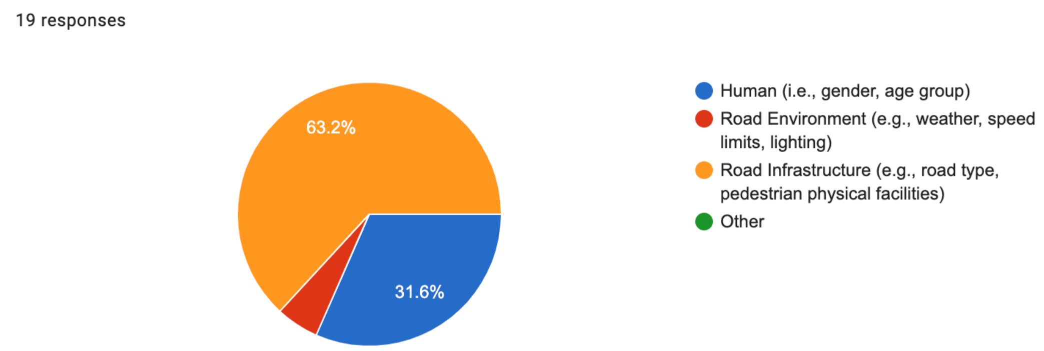


Figure 9.22.: Road-crash priority factors to improve pedestrian safety

In Figure 9.22, the responses prioritised road factors to improve pedestrian safety as follows: infrastructure (62.3%), human (31.6%), and environment (5.3%) [see Section 8.3]. By contrast, the prototype ranked four highest attributes that affect road-crash casualty rates under the following categories: environment (i.e., fine weather with no wind; speed limits set at 30 mph), infrastructure (i.e., absent pedestrian facilities), and human (i.e., male). However, compared with actual data ahead in time, on environmental factors the prototype indicates opposite trends. More particularly, the weather and the speed limits presented an underperforming MOE in terms of accuracy. Furthermore, regarding the infrastructure, at no pedestrian physical facility within 50 m [see Figures 8.10, 9.21], it was not possible to compare the prototype projected trends due to lack of data from 2018. Lastly, regarding human factors these demonstrated to align the actual data ahead in time, by gender. More in detail, the prototype indicated males with the highest tendency of exposure to road fatalities. Moreover, an increasing tendency of female exposure to road-crashes was indicated [see Section 8.3].

Figure 9.23 illustrates some of the contributions of the participants to address human error factors that affect pedestrian safety.

Can you **explain** ways you have for addressing both human error factors affecting pedestrian safety? Give some examples of concrete effective measures(e.g., by gender and age group).

12 responses

Human error can be addressed by taking a Safe System approach, and providing a forgiving road environment that minimizes the chance of human error having a fatal or serious outcome. For example, providing a slow speed environment, supported by infrastructure (e.g. traffic calming) if an error occurs, it will not result in a fatal outcome.

Increasing the visibility of young and elderly pedestrians during night-time conditions using reflective clothing.

To adopt a human-centric infrastructure where we put pedestrian safety as a priority. It can be done through traffic calming

pedestrian fencing installation on "complex" pedestrian crossing situations to minimize the possibility of error and prevent pedestrians walking on the pavement - good both for young and old. Good illumination in underpasses and elevators in underpasses - especially praised by older generations

Activities related to the knowledge, behaviour and attitudes change

Separate vulnerable road users facilities, separated at level cross-passing

Figure 9.23.: Contributions to address pedestrian safety due to human error

Figure 9.23 illustrates an overview of the contributions of the participants for addressing human errors that affect pedestrian safety. Specific comments included on words '...visibility of young and elderly pedestrians during night-time...', '...pedestrian safety as a priority', 'Separate vulnerable road users facilities...' and '...training and retraining...'. No comments on how to differentiate training by gender were found [see Chapter 4, Section 8.3.1].

Figure 9.24 presents the ranking of the participants in environmental conditions that affect pedestrian safety.

6. In terms of the road environment, the findings indicate that adverse environmental conditions (e.g., weather), speed limits, and lighting systems are major influencing factors in pedestrian road-crash fatalities. How confident are you that these environmental conditions affect pedestrian safety?

17 responses

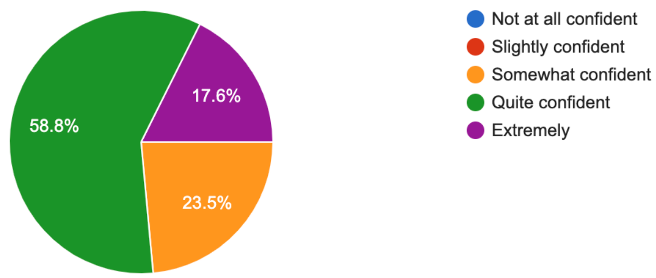


Figure 9.24.: Environmental conditions affecting pedestrian safety

In Figure 9.24, most responses of the ranking regarding environmental conditions (i.e., weather, speed limits and lighting conditions) affecting pedestrian safety indicated to be quite confident (58.8%), somewhat confident (23.5%) and extremely confident (17.6%). By contrast, the findings of the prototype by the weather mostly fails to project the trends. Moreover, by speed limits and by lighting conditions the prototype presented underperforming accuracy measures [see Section 8.3].

Figure 9.25 presents some of the contributions from the participants regarding preventive measures to address environmental conditions affecting pedestrian safety.

7. Can you **provide** with recommendations with concrete examples, for reducing the impact of road environmental conditions such as rain and wind, permanent speed limits set at 30mph, and lower levels of lighting that result in pedestrian road-crash casualties?

13 responses

good drainage of the infrastructure, signing and lighting, pedestrians facilities , speed limit around residential areas

improved lighting (directed); road bumps

I don’t think rain and/or wind affects too much in Spain to these crashes. Only the speed limits should be into consideration

Installation of speed limit device in all the categories of vehicles

at the moment, the only thing that comes to my mind is the regular maintenance of drainage canals. In my town, drainages are oftentimes clogged by dirt, which prevents rain water to drain

To adopt a human-centric infrastructure where we put pedestrian safety as a priority. It can be done through traffic calming

Figure 9.25.: Road environment affecting pedestrian safety

Figure 9.25 presents some of the contributions of the participants preventive measures to address environmental conditions that affect pedestrian safety. Specific answers included on words ’speed limit around residential areas...’ ’ I don’t think that rain and/or wind affect too much in Spain...only speed limits should be taken into consideration...’, ’...pedestrian safety as a priority’, ’In my town, drainages are oftentimes clogged by dirt...’ and ’...training and retraining...’. No comments on how to implement pedestrian safety facilities due to adverse weather (e.g., snow) were found [see Chapter 5, Section 2.4.2].

Figure 9.26 presents the ranking of the participants regarding the absence of pedestrian physical facilities affecting pedestrian safety.

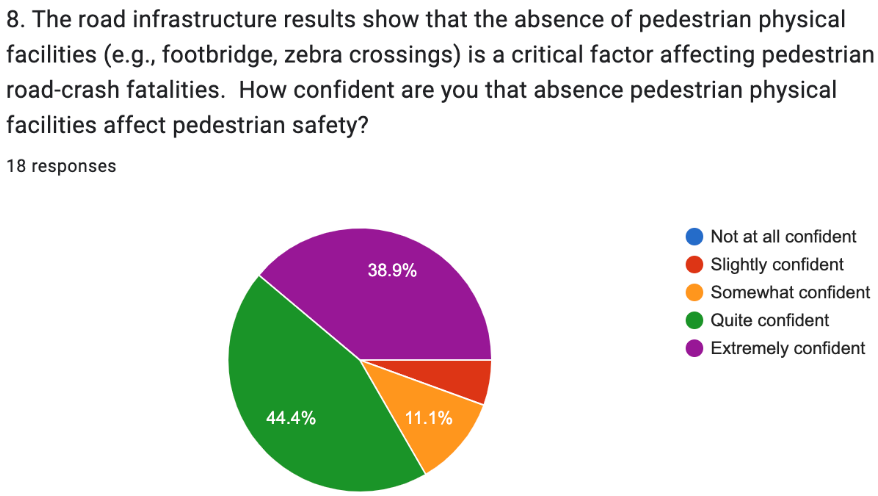


Figure 9.26.: Absent pedestrian physical facilities affecting pedestrian safety

In Figure 9.25, most of the ranking responses regarding the absence pedestrian physical facilities that affect pedestrian safety indicated to be quite confident (44.4%), extremely confident (38.9%), somewhat confident (11.1%) and slightly confident (5.6%). In contrast, the prototype indicated an increasing tendency for most types of pedestrian physical facilities except for absent pedestrian physical facilities where a decreasing tendency was projected. However, due to the lack of available data from 2018 on the classified pedestrian physical facilities (e.g., absent physical facilities, pelican) the findings cannot be compared [see Section 8.3.3].

Figure 9.27 presents the contributions of the participants in addressing the existing challenges regarding absent pedestrian physical facilities affecting pedestrian safety.

9. Can you **provide** with concrete measures that can mitigate the impact of absent pedestrian physical facilities leading to pedestrian road-crash fatalities?

12 responses

Increase investments in pedestrian safety infrastructure.

sidewalk/footpath, crosswalk, refuge islands, traffic calming, lighting, assessable, dedicated bike lane, awareness/education

Absence of adequate pedestrian facilities can contribute significantly to pedestrian road-crash fatalities. To mitigate this, it's crucial to invest in the development of pedestrian infrastructure. This includes constructing sidewalks where there are none, ensuring they're wide enough for safe pedestrian traffic, and are free from obstructions. Creating well-marked, visible crosswalks and pedestrian bridges or underpasses in areas with heavy traffic can also provide safe crossing options. Introducing pedestrian refuge islands on wider roads gives individuals a safe place to stop if they can't cross the entire street at once. Additionally, installing pedestrian signals at intersections and implementing traffic calming measures such as speed bumps or roundabouts can help slow down vehicles and give priority to pedestrians.

Our research shows that this is critical. Providing a footpath that is free from obstructions, as well as appropriate crossing facilities can almost eliminate pedestrian fatal and serious outcomes. Crossing facilities should include raised crossings. The evidence regarding the use of footbridges is not good, as these are seen as inconvenient in many countries, and so are not used.

Figure 9.27.: Challenges regarding absent pedestrian physical facilities

Figure 9.27 presents some of the contributions of the participants preventive measures to address absent pedestrian physical facilities affecting pedestrian safety. Specific answers included sentences such as '... Creating .... visible crosswalks and pedestrian bridges or underpasses...', '...the evidence regarding the use of bridge is not good, as these are seen as inconvenient in many countries...', 'To adopt a human-centric infrastructure ...'. No comments on how shared spaces could be adopted to improve pedestrian safety were found [see Chapter 5, Section 2.4.5].

Figure 9.28 presents some of the suggestions for future research by the participants regarding cognitive disabilities that affect pedestrian road crashes.

1. The findings of this show a paucity of literature on various levels of cognitive disabilities affecting pedestrian road crashes. What would you **suggest** to prepare for future research on this topic ?

11 responses

I'm not sure.

Establish collaborations with relevant stakeholders

To facilitate future research into the impact of cognitive disabilities on pedestrian road crashes, it is essential to adopt an interdisciplinary approach, leveraging expertise from fields like transportation engineering, traffic psychology, special education, occupational therapy, and neurology. Data collection should involve building databases documenting instances of pedestrian accidents involving individuals with cognitive disabilities, potentially through collaborations with healthcare and special education institutions. The design of these studies should be inclusive, integrating the experiences of people with cognitive disabilities via interviews, surveys, and focus groups. Research should also employ simulations and observational studies to understand behavior in controlled and real-world environments.

This is an important topic, especially reduced cognitive abilities for older road users. When older road users are walking, they are very vulnerable to injury. However, I think the greatest benefits will come from making decision makers more aware about the solutions we have now that are already available and proven to have an impact. These will have benefits for those with cognitive disabilities. Put another way, we don't need to wait for this type of research before we act.

Figure 9.28.: Pedestrian cognitive disabilities impacting on pedestrian safety

Figure 9.28 presents some of the contributions the participants on future research to improve pedestrian safety regarding cognitive disabilities. Some of the answers from the participants included '...we don't need to wait for this type of research before we act...', 'I'm not sure.', 'Tech companies have speed data that can help a lot...'. No comments on suggesting further research to be adopted in different regions on this topic were found [see Chapter 4].

Figure 9.29 presents recommendations of the participants for improving data collection procedures.

Can you **provide** with concrete recommendations to address the data collection procedure ?

12 responses

Provide a very easy app for traffic agents connected with a BIG centralised database which would be capable of alerting of fast targeted Road safety inspections and other Road safety measures

Easy accessibility to data collection from contributory elements

Drone footage, action camera footage, and many more, all of this can be processed by AI softwares

This can be overcome using multi-sources of data including data generated from computer vision and image processing. AI will help in finding the pattern and generate insights from that patterns.

comprehensive data about road

It's a coincidence, I just made a recommendation in my example above :). But there are many opportunities with historical speed data. I'm involved in a new idea to find areas on the road where the recommended speed limit is dangerous. I would be happy to share this with you on a call or something.

I'm not sure.

Figure 9.29.: The improvement of data collection procedures

Figure 9.29 presents some of the contributions of the participants in improving data collection procedures. Some of the answers from the participants included, '...multi agency collaborative to enhance data standard and quality.' 'We do lack data on the impact of pedestrian exposure...', '...better coordination between main road safety stakeholders', '...data generated from computer vision and image processing. AI will help in finding the pattern and generate insights from patterns...'. No comments on how to improve the procedure of data collection were found [see Chapters 4, 5, and 7].

Figure 9.30 presents the ranking with views of the participants on the applicability of adopting AI to examine pedestrian safety with a similar approach in different regions.

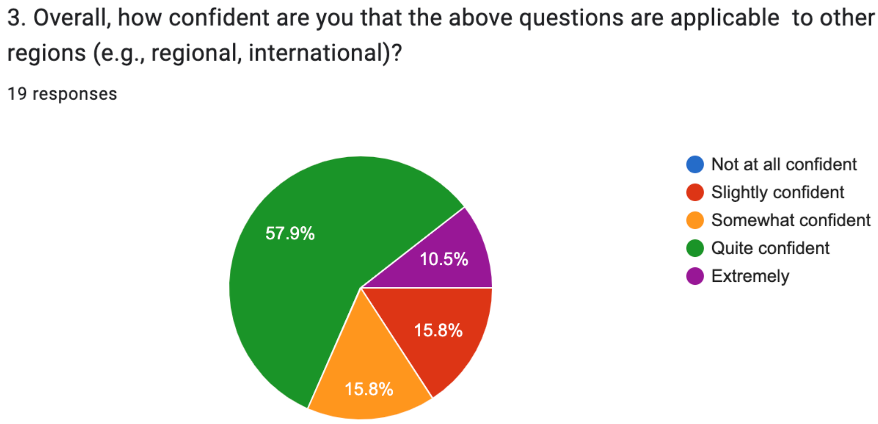


Figure 9.30.: The applicability of AI cross-region to improve pedestrian safety

In Figure 9.30, many of the responses on the applicability of adopting AI to examine pedestrian safety in different regions, indicated that they were quite confident (57.9%), somewhat confident (15.8%), slightly confident (15. 8%) and extremely confident (10.5%). By contrast, the prototype demonstrates its applicability to different regions [see Section 8.2.1].

9.4.4. Lessons Learnt: Expert Opinion

As part of this study, the deployed questionnaire to a designated group of engineers (i.e., focus group), presented an opportunity to comprehend additional ways in which the knowledge and experience from international professionals can contribute to improve pedestrian safety. Even though the focus group is not representative of all engineers worldwide, the expert opinion collected, can provide a foundation for the development of new approach that successfully meets the needs to improve pedestrian safety. However, the findings of this questionnaire did not point to topics related to improvement procedures of data collection such as standardisation [see Chapters 2, 7, Sections 2.5.2, and 10.3]. Another aspect that is not present in the answers of the questionnaire, regarded the direction of future research on pedestrian safety approaches for pedestrians with cognitive disabilities. Disappointingly, for pedestrians with cognitive disabilities (e.g., dementia), the prototype does not project any trends. This is because the database (i.e., STATS19) does not include that type of details on the metadata (e.g., pedestrians with dementia by gender, location, level of disability) [see Sections 2.3, and 6.2.1]. On the other hand, even though a high rate of non-recorded ages were found, for the elderly, the prototype still indicates an increasing tendency which aligns with the actual data [see Chapters 2, and 5] and 8.3]. In summary, the results of the questionnaire carried out in this study, are significant in at least two major respects: developing new approaches and raising the awareness of the participants. This is particularly valuable to the prototype deployed in this study. Overall, the questionnaire was well received and more importantly for further progress, it confirmed the willingness of the participants to support improvements in pedestrian safety. As an example, some of the participants reiterated their appreciation to "contribute with new ways" to improve pedestrian safety. Moreover, some participants proactively offered for future collaboration. To the success of these initiatives, further investigations to establish how to engage a wider community at an earlier stage are needed.

9.5. Observations

Overall, the data analysis conducted in Section 9.3 demonstrates its applicability for short-term projections, specifically one year ahead. However, the results of the sensitivity responsiveness in this chapter reveal limitations for describing future events. Similarly, the findings from the expert opinion demonstrate similar constraints. A possible explanation for this might be a lack of data quality, which has been emphasised by the deployed prototype. Nevertheless, an interesting finding that emerges from systematically correlating data about pedestrian general attributes (Uzan and Wagstaff, 2017; Hezaveh and Cherry, 2018) and the road infrastructure (Sullivan and Flannagan, 2002), with expert opinion, is that these can complementarily contribute to deepening the understanding of previously unexplored patterns leading to road crash-casualties [see Chapters 2, 4, 5, and Section 9.4]. For future research work, the provision of standardised interfaces to integrate the data should be considered. This is a crucial measure to help prevent interoperability issues.

9.6. Extracted Knowledge on Pedestrian Safety: Practical Applicability

This study confirms that implementing relevant pedestrian physical facilities, can considerably reduce pedestrian safety road crashes. Moreover, consistent with the literature, the absence of pedestrian physical infrastructure, significantly worsens pedestrian safety [see Chapters 2, 5, 8, and 9]. These results, support the idea of re-assessing the influence of each type of the existing pedestrian physical facilities. Additionally, they support to recommend targeted training that can prioritise different levels of road-crash exposure (e.g., male, elderly, and children). Moreover, they suggest exploring methods that can measure different levels of cognitive disabilities. Furthermore, to introduce physical facilities accounting for weather condition during the day, for illuminance at night, and for locations with permanent speed limits set at 30 mph. However, the main weakness of this approach to be successful is the lack of data quality [see Section 8.3]. An implication of this finding is that it might affect the system reliability. Therefore, it must be interpreted highly cautiously.

10. Discussion

10.1. Introduction

The adoption of ML to systematically investigate pedestrian safety has proven its applicability and complexity. The latter happens due to the influence of various road elements, adding a significant level of unpredictability (e.g., human behaviour, weather). Therefore, a pre-defined characterisation including an inter-relational, prioritised, and streamlined working framework, is vital. This chapter discusses the progress made with the FCM prototype deployment [see Chapter 8] under the following headings:

10.2 Contributory Factors

10.3 The role of Data Quality

10.4 Prototype Deployment

10.5 Systemic Data Analysis and Evaluation

10.6 Data Quality Improvement

10.2. Contributory Factors

The first phase entailed investigating a number of contributory factors influencing pedestrian-road crash casualty causation. During this phase, the methodologies currently used were further explored. Subsequently, relevant information was selected, followed by its incorporation into the deployed prototype. The information was categorised into three main categories: pedestrian general attributes, road environment, and road infrastructure.

10.2.1. Pedestrian General Attributes

Looking into the pedestrian general attributes in a road environment, pedestrian safety models by gender (i.e., male, female) and age group (children, young adolescents, adults, elderly) were previously explored [see Section 2.3.1]. The results show the elderly, children, and females as the most vulnerable. However, interestingly, overall, males were the ones accounting for the highest exposure to pedestrian road-crash casualties. Furthermore, a note of caution is needed for pedestrians diagnosed with cognitive disabilities since a limited body of both quantitative and qualitative literature was found. These findings propose scrutinising information on pedestrian general attributes in more depth. However, pedestrian safety models are highly dependent on the quality of their reliability and consistency (Hatfield and Murphy, 2007; Twisk et al., 2015; Nemire et al., 2016; Earl et al., 2018). Therefore, adopting more sophisticated pedestrian safety models to improve data quality while reducing data dependency is highly recommended.

10.2.2. The Infrastructural Urban Road Environment

The urban road environment explored two different groups of pedestrian safety models: road environment (i.e., weather, speed limits, lighting), and road infrastructure (i.e., road type, pedestrian physical facilities). As can be seen in Section 10.2.1, a significant number of pedestrian safety models demonstrate a high dependency on the quality of the data.

Road Environment

Regarding the forecasting accuracy of the FCM prototype (i.e., descriptor), the weather seems the most challenging to describe. Nevertheless, evidence collected on weather conditions influencing road-crashes in GB indicates a general least likelihood of pedestrian road casualties when raining. In addition to that, a study conducted in Greece using a 21-year dataset, indicated increases in rainfall reduce road-crash fatality trends as physical activity levels tend to decrease [see Section 2.4.2]. In the same vein, another useful study shows pedestrians perceiving the weather as a barrier to physical movement (e.g., poor visibility) (Mohamed et al., 2013). Regarding the effects of permanent speed limits in a road environment, the adjustment of speed limits (e.g., introducing variable speed limits at rush hour) and the adoption of traffic calming treatments have been previously proposed [see Section 2.4.3]. As regards lighting, the results prove higher illuminance ratios reassure the pedestrians cognitive judgements (e.g., detecting obstacles). However, in this context, the MAPE examined in this study was found to be limited to the location of experimentation [see Section 5.2.3]. Therefore, the adoption of this method as complementary or the use of more sophisticated models requiring a lower amount of data is proposed [see Sections 2.4.4 and 5.2.3].

Road Physical Facilities

In this group, pedestrian safety models considered two main groups: intersections and roundabouts. Regarding intersections, overall, T or staggered junctions represented the highest pedestrian road-crashes due to poor sight at multiple conflict points. On the other hand, roundabouts and mini-roundabouts demonstrated their potential to reduce pedestrian road-crashes by up to 30% (Li et al., 2012; DfT, 2018; Vignali et al., 2020). However, surprisingly, a parameter or feature that differentiates between unsignalised and signalised intersections was not found. This is rather disappointing, as according to many, each type of sign at an intersection regulates different traffic characteristics (Prasetijo and Ahmad, 2012; Feliciani et al., 2017). For example, should the unsignalised intersection account for the geometric design based on the number of pedestrians crossing, its applicability would become far more useful. Therefore, should a signalised intersection be designed as part of a traffic system based on the pedestrians' crossing speeds, the same principle should apply (Onelcin and Alver, 2017). As regards pedestrian physical facilities, although the analysis did not directly focus on the effects of present vs. absent pedestrian physical facilities, the presence of pedestrian facilities was indicated as a fundamental factor in reducing pedestrian road-crash casualties. For instance, in GB, most road-crashes occurred when pedestrian facilities were absent (67.6%) followed by pedestrians located at pelican or similar (14.1%) (DfT, 2018) [see Chapter 8]. The selected pedestrian safety models for examination were for the zebra, since, as explained in Section 5.3 pedestrian safety models for the pelican, footbridge and private property were not found. About the zebra, according to Al Bargi and Daniel, 2020 a key finding was that pedestrian crossings seem not to be directly influenced by the design of the road infrastructure. Chapters 2, 4, and 5 agree.

10.3. The role of Data Quality

The second phase of the prototype deployment consisted of collecting data from a database (DfT, 2017; DfT, 2018). The evidence indicated data quality to play a fundamental role to enable a systematic investigation to improve pedestrian safety [see Section 6.5]. Data mining was adopted to systematically identify previously unforeseen patterns by applying the FCM on road data that combined pedestrian general attributes and road elements. The undertaken analysis demonstrates that indicative trends can be concluded based on the attributes under examination. Different scenario rules (e.g., gender: 75%; pelican: 85%) present different levels of accuracy rates. In addition to that, when it comes to projecting unforeseen events, the lack of data (e.g., NRA), may lead to misleadin conclusions of the system based about the analysed road elements (in this case, the age group) and the entire analysis. Therefore, enhancing the consistency of databases (e.g., content, algorithms, set of rules) to improve data quality is crucial [see Chapters 6, 7]. This is because data quality can ensure scalability to multiple regions while facilitating the development of a policy, standard, or procedure [see Section 2.6]. Nevertheless, much debate on the benefits of using datamining over traditional statistical analysis and vice versa has been carried out [see Chapter 2]. Each form of analysis has certain advantages and shortcomings. A comparative analysis between traditional statistics and ML follows.

10.3.1. Statistical Analysis: Data Samples

A Traditional Statistical Analysis

For approximately a decade (2009-2019), in GB, road data reported to police showed an overall average decrease in pedestrian road-crash casualties, where a steep decrease of 20% was noticed. Moreover, it showed a continuous decreasing tendency of pedestrian road-crash casualties' tendency between 2009 and 2015 (DfT, 2015). On the other hand, between 2015 and 2019, some fluctuations were noticed. Table 10.1 presents the pedestrian road-crash casualty over a 5-year period (2015-2019).

Table 10.1.: Pedestrian road-crash casualty data: 2015-2019

Year	Trend
2014-2015	$\downarrow$ a drop by 8% (DfT, 2015)
2015-2016	$\uparrow$ a marked increase by 10% (DfT, 2016)
2016-2017	$\nearrow$ a gradual increase by 5% (DfT, 2017)
2017-2018	$\swarrow$ a slight fall by 3% (DfT, 2018)
2018-2019	$\rightarrow$ remained steady with 3% (DfT, 2019)

As can be observed in Table 10.1 two marking periods are noted: 2015-2016; 2016-2017. When comparing both periods, a 50% reduction in the overall number of pedestrian road-crash fatalities is noticed. Even so, the first period shows a 10% decrease compared to the previous year, and the second period shows a 5% increase. The first period (i.e., 2015-2016), shows a steady increase from 408 in 2015 to 448 in 2016. Additionally, on the pedestrians' gender, 56% were reported as males, whereas 26% of the age group was aged between one and 15 years old. The records also show that 36% of pedestrian road-crashes were registered between 3 p.m. and

7 p.m. Regarding the road type classification, the *in built up roads* (i.e., roads with permanent speed limits at 40 mph accounted for 18% of the pedestrian road-crash casualty, compared with 2% on other type of roads (e.g., non-built up roads, motorways) (DfT, 2016). The second period (i.e., 2016-2017), shows a gradual increase from 448 in 2015 to 470 in 2016. Like the first period findings, the gender of the pedestrians remained, whereas in the age group between one and 15 years old, only a slight decrease (i.e., 1%) was noticed. Regarding the road type classification, urban roads involved a higher number of pedestrian road collisions (i.e., 20%) compared with 5% on rural roads. One interesting finding was that, compared with the first period, the terminologies for the road classification were modified (or readjusted from *in built up roads* to rural roads and from *non-built up roads* to urban roads) (DfT, 2017).

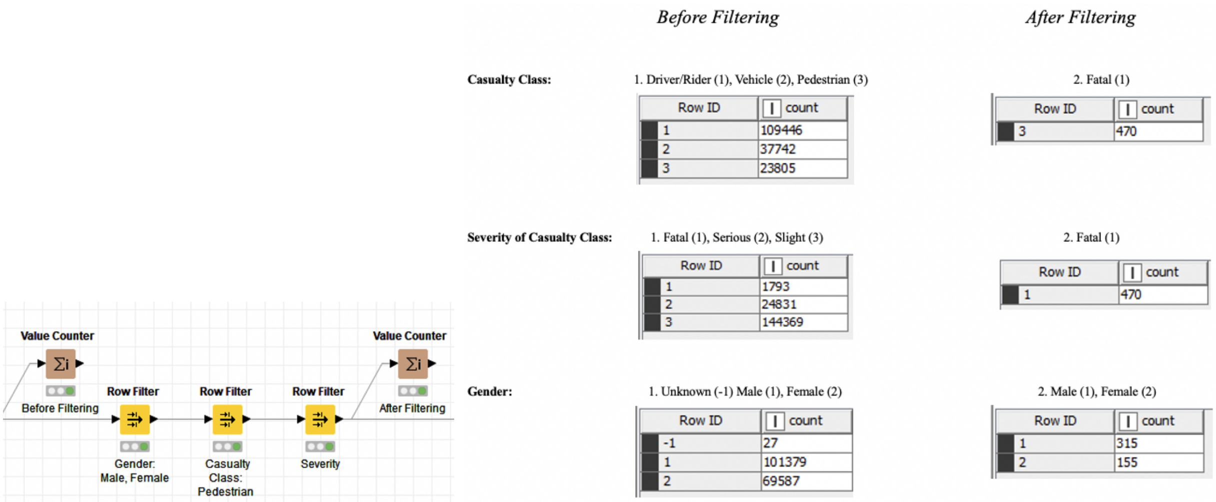
A Statistical Analysis using KNIME

The results obtained from the 10-year period data analysis conducted in this research align with findings from official sources. Figure 10.1, illustrates the extracted data from the pre-defined requirements.

Upon exploring the data, Figure 10.1b shows out of 23.805 accidents involving pedestrians, 470 resulted in pedestrian fatalities, of which 315 were males and 155 were females. These results align with the official sources (DfT, 2017). However, some other expected limitations of this approach are as follows:

i. Filling out the actual road data form manually: possible misspelling and misunderstanding of the information when transferred into a database.

ii. Filling missing fields manually on the database: misunderstandings or a higher



Row ID	count
1	109446
2	37742
3	23805

Row ID	count
3	470

Row ID	count
1	1793
2	24831
3	144369

Row ID	count
1	470

Row ID	count
-1	27
1	101379
2	69587

Row ID	count
1	315
2	155

(a) Data filtering process (b) Data extracted results

Figure 10.1.: Extracted records

time consumption in processing the data as handling large sets of data (e.g., 23.805 accidents) (Figure 10.1b).

iii. Discarding unknown records: discarding all the unknown values; for example, 27 accidents with the unknown gender were discarded (Figure 10.1b). This is a key issue that can directly affect the analysis's applicability to reality.

iv. Using the mean: The mean of each attribute contains missing values to fill these fields. This only applies to databases consisting of numerical values.

Despite the limitations mentioned above, the short-term data projections undertaken in this study surpass standard expectations by achieving an acceptable overall accuracy rate of 97.03% [see Section 8.2.2, Table 8.1]. Additionally, for each of the examined factors, it establishes three distinct rules to test the model response at 80%, 75%, and 85% respectively [see Sections 8.3, and 9.3]. Furthermore, this section introduces how the prototype can be applied to the entire database for achieving long-term data projections (i.e., over a 10-year period).

10.3.2. Computational Datamining Tool Selection: KNIME

A variety of ML pedestrian safety models have been deployed to explore the influence of pedestrian general attributes (i.e., gender, age group), the road road environment (e.g., weather, speed limits, lighting) and the road infrastructure (e.g., road type, pedestrian facilities) on pedestrian road-crashes [see Chapters 4, 5, Section 7.3]. The FCM prototype was deployed on KNIME a software application tool [KNIME v 3.7.2, 2019]. As previously introduced in Section 3.3. KNIME is a computational datamining tool that integrates various components for modelling algorithms. It uses modular data flows, modules or pipelining "Lego of Analytics" concept (KNIME, 2021) [see Sections 6.2, and 6.2.1]. KNIME integrates various other open-source projects (e.g., ML Weka, H2O.ai, Keras, Spark). This means that depending on the data, KNIME offers options of the most appropriate programming and configuring language[s), not just to the application but also to each part of the application (KNIME version 3.7.2, Release date: April 18, 2019). These features are particularly useful since the deployed system account for communicating smoothly with the different interfaces (e.g., native connectors, data transformation, visualisation, export). In summary, KNIME came across as promising tool to assess the effectiveness of the FCM applicability on examine road data. This because its in-built capabilities demonstrated relevant capabilities to deploy the required system with a considerable level of flexibility.

10.4. Prototype Deployment

This section regards the FCM prototype deployment features using KNIME. With respect to the system capabilities, processes and techniques of exploring data are explicitly described in Chapter 6 (Ehsaei and Evdorides, 2011; Massaro et al., 2020).

Data exploration regarded two main types functionalities: data configuration, data transformation [see Section 6.2]. A few examples are shared below:

Data Configuration: with respect to data filtering, Table 3.1 shows the selection procedure (e.g., Car_passenger added to excluded variables, Casualty_Class added to included variables) of the data configuration functionality [see Section 10.3.2].

Data Configuration: in terms of data verification, Figure 3.7 shows the data split into two sub-modes: descriptor (e.g., 80%), predictor (e.g., 20%). This means the descriptor refers to 80% of trained data, and the predictor to 20% of tested data. Both modes were verified on the FCM demonstrator. The descriptor sub-mode was explored on the FCM prototype, and tested the use of new rules FCM sensitivity response [see Section 10.3.2].

Data Transformation: on data association, Figure 6.3 illustrates a joiner which is a data transformation functionality [see Section 10.3.2].

In Chapter 6 Figure 6.6 shows that after selecting an ML algorithm, the quality settings enabled to statistically score the inter-relational values between the selected attributes. This is categorised as a data transformation functionality [see Section 10.3.2]. Regarding data visualisation, the format .svg was used. To the proposed analysis, included data verification (i.e., descriptor 50%, and predictor 50%), data analysis (i.e., descriptor 80%), and data validation with a set of new rules (i.e., descriptor 75%, and descriptor to 85%). Although this systematic investigation enabled considerable flexibility, there are no guarantees to obtain the all the possible answers as these may still result in lacking functionality and also efficiency (e.g., data correlations) [see Chapter 9, Section 6.4.2]. Summarising, the representational pipelining provided by KNIME, enables formulating problem-solving techniques described with a series of data flows and relationships, expressed as modules. The following sections present the adopted FCM applicability more in detail.

10.4.1. FCM: GB Pedestrian Road-Crash Casualty Causation

This section describes the three main steps taken to deploy the FCM, by using using actual pedestrian road-crash data from the GB in 2017 (DfT, 2017) [see Chapter 8]. The knowledge discovery approach on GB pedestrian road-crash data was undertaken as follows:

i. Statistics: GB [see Section 10.3.1].

ii. FCM demonstrator: GB vs. London (i.e., statistics, descriptor, predictor) [see Sections 8.2.2, 8.2.3].

iii. FCM prototype: GB (i.e., statistics, descriptor) [see Section 8.3].

iv. FCM sensitivity response: (i.e., statistics, descriptor: new rules) [see Chapter 9].

The data was made available both from a public and official source (DfT, 2017; DfT, 2018). Additionally, to visualise the statistical road data geographically, a tool called Google Earth Education was used [see Section 8.2]. The actual road data was provided by the public data source STATS19, which was selected (e.g., pedestrian attributes, road elements) and combined using KNIME, in all forms. Under these conditions, the FCM was applied to actual road data from 2017. The discovered knowledge from the FCM was then compared with actual road data from the following year (i.e., 2018). However, it was felt that the gathered knowledge could have been provided with a higher degree of reliability. This because, whilst the prototype was being deployed, some of the findings triggered the need to further investigate a few distinctive characteristics. An example was that the investigated data during a 1-year period described 10 pedestrian fatalities on a single accident which accounted for 3.42% of the total number of pedestrian fatalities. Similar results from the official sources were not found [see Section 8.2.2]. Even though being out of the scope of this study, the findings suggest further investigation on other types of road users (e.g., drivers) and road elements (e.g., vehicles). In summary, the objectives to discover knowledge by incorporating road data into KNIME and applying the FCM was successfully achieved. Therefore, KNIME proves the applicability of modules and KDD. This to both manage data and extract relevant knowledge in a consistent and efficient manner [see Sections 3.4, 10.3.2]. However, the prototype only proves to be study-specific as it does not provide reasoning capabilities for different sets of data [see Chapter 8; Section 8.2]. To develop a comprehensive analysis, the following requirements were assessed:

Criteria specification: based on a number of assumptions combined with a range of variables to enable getting a clear representation of the examined data using KNIME.

Applicability into actual data: to verify and demonstrate the applicability of the mechanisms used for discovering new knowledge using the FCM.

Sensitiveness testing: to both create and formulate new rules in a simple and easy manner.

Although the FCM is understood as an effective concept based on a set of procedures (e.g., rules, techniques), from the examined road elements (i.e., pedestrian general attributes, road environment, road infrastructure), it was not possible to exclusively rely on the extracted knowledge to draw firm conclusions on reasons underpinning pedestrian road-crash casualties in an urban road environment. Still, recent advancements show other computational tools applying more sophisticated unsupervised ML models in real-time and also to different regions [see Section 7.2.4]. Moreover, the latest deployed tools may likely provide a more user-friendly and more clarified procedural development (e.g., new features to the algorithms, aggregation of the modules). Nevertheless, the tests conducted using the FCM prototype and the FCM sensitivity response were successful. In addition to that, the results showed some degree of alignment between the described trends and the actual data (e.g., gender, lighting with a chance of a fatality at night). Moreover, the trends objectively point to relevant rules that are applicable to specific factors [see Sections 8.5, 9.5]. However, based on this specific study, understanding how external factors (i.e., computational capabilities, pedestrian general attributes, and road elements) can influence the raw data to explore new knowledge lies in the unknown.

10.5. Systemic Data Analysis Evaluation

KNIME was adopted to handle several data exploring procedures and apply ML algorithms into road data [see Sections 6.2, 6.3]. The exercise included tasks as follows:

i. Data pre-processing.

ii. Classification (e.g., clustering, regression).

iii. Data verification.

iv. Visualisation.

v. Feature selection, configuration.

Regarding the computational datamining tool term "data analysis evaluation", any data analysis entails two fundamental aspects: verification and validation. Within these two terms characteristics include the following items:

1. **Verification**

a) Objectivity: identifies the applicability of the proposed methodology to analyse the data.

b) Applicability: qualifies the evaluation of the established procedure to examine the data.

c) Consistency: provides knowledge consistently when applying the same model, to the same set of data, under the same circumstances.

d) Maintainability: provides knowledge consistently when using different sets of data, and also when using either the same or different algorithms under the same circumstances.

2. Validation

a) Verification: reproduces sample data the same way, in the same circumstances.

b) Sensitivity: assesses the sensitivity response, with different ratio data in the same way in the same circumstances.

c) Reliability: ensures consistent and reliable knowledge with different sets of data, and either the same or different algorithms in the same circumstances.

d) Costs: analyses and interprets the costs in the same way in the same circumstances.

e) Certifiability: ensures the correctness of the processes implementation, both in manner and context of operation.

In this study, the evaluation process was simplified by investigating the system applicability in the form of a prototype. On the prototype assessment, statistical significance was analysed using several verification testing techniques as appropriate. However, to enhance its accuracy it may be the case that the validity of the findings requires using more sophisticated methods. Moreover, it was felt that for this study, maintainability, reliability, costs, certifiability were more related to evaluate a fully working system. Therefore, these were not included on the data analysis evaluation.

10.5.1. Systemic Data Analysis Evaluation: Observations

Although the FCM does not fully demonstrate the initially aspired competences, it has been clearly noticed a major step forward. Therefore, a few observations of the FCM were formulated as follows:

1. The verification demonstrates the prototype as objective, applicable, and consistent. This since, based on the proposed procedures under the same circumstances, it provides consistent knowledge. Moreover, this is ensured by executing relatively simple procedures and for the clarity that KNIME offers by using the modular data flow [see Section 6.2].

2. The validation shows he prototype to be highly dependent on data quality. Moreover, depending on the scenario rules set by the ML algorithm, the accuracy rates vary. More information can be found in Chapter 9, Sections 3.3.3, and 6.3.

The deductive mechanisms used for the selection criteria of the road elements as for example junction location by region, is out of the scope of this study. Therefore, assessing maintainability, reliability, costs, certifiability would become quite likely challenging. Moreover, the complexity of the subject made clear existing disadvantages of combining multiple algorithms in the present prototype. This because of, among others, its limitations to consider alternative rules at the same time. Nevertheless, further improvements on the FCM are suggested, as to develop a clarification framework on the objective's requirements. Respectively, allow for characteristics as maintainability to the system verification, reliability, costs, and certifiability to the system validation. Furthermore, research associating more than one algorithm to formulate a prototype is highly suggested.

Performance of the System

As regards the performance of the system in terms of consistency, based on the provided features by the individual modules the overall and detailed system observations [see Chapters 8, and 9] it may be concluded that:

1. The selection of the appropriate application feature such as "use random seed" values for the data verification indicates a challenge when manipulating data. This data should be more intelligently stored, accessed, updated, and selected. This because, depending on the type of the analysis, other applications than KNIME, may be in position to offer better approaches for the operation.

2. The feature prunning reduces the complexity of the final classifier, while improving the accuracy by the reduction of overfitting. The feature gini measures statistical dispersion of the module. To deploy the features for road environment (e.g., weather), as these should be addressed more efficiently and clearly than KNIME currently does.

3. As the results were not fully encouraging, integrating a database, a knowledge assessment base and an analytical tool into a full compatible unified system is proposed. However, unfortunately it has been found that the available software packages share limited compatibility between each of the presented features. Therefore, any further attempt to create a final operative system should take this critical aspect into account.

10.5.2. Systemic Data Analysis Evaluation: Summary

In a wide range of scientific and engineering domains, the FCM has been at the core of the conducted study [see Chapters 2, 7, 8, 9]. The investigation of FCM has shown that, the expectations of what may be achieved have been raised to a significantly higher level. Since the study was limited to commercially available software, its potential uses remain limited to several factors. Notwithstanding these limitations, the study proves its capabilities for interpreting relational links between humans and the road elements leading to pedestrian road-crash casualties, such as for example involving pedestrians with a level of physical disability [see Chapters 2, and 4]. Additionally, the road infrastructure highlighted the lack of pedestrian safety models on a number of pedestrian physical facilities (i.e., pelican, footbridge, private property) [see Chapters 2, and 5]. Despite the limitations of the research conducted, it certainly adds to our understanding that to provide a reliable road data analysis when using the FCM applied on road data, the quality of the data is vital [see Chapter 6]. One source of weakness of KNIME was the modular application (e.g., modular analysis, software, hardware) proving to have influenced the performance during this study [see Chapter 6, Section 10.5.1]. Nevertheless, in overall it was felt that the observations made, will be helpful in the future development of an operating system.

10.6. Data Quality Improvement

This study was carried out to confirm that, ML can be a proactive contribution to examine road data and identify unforeseen links affecting pedestrian safety. Although the findings demonstrated that some characteristics play a critical role in determining data quality, the lack of data quality lays a critical issue. Since data are a vital component to leverage data-driven decisions, bellow is briefly described some of the suggestions that can be introduced to improve data quality by using ML algorithms.

a) Accuracy More sophisticated ML algorithms. Some of the advantages of adopting progressive ML algorithms and systems:

[a.1.] To improve operating with a lower dependency of historical data and with different formats (e.g., structured vs. unstructured).

[a.2.] To enhance the capabilities of a singular analysis that can, incorporate both qualitative and quantitative data in near real-time.

[a.3.] To improve the level of the accuracy, and to evaluate the systems viability.

[a.4.] To identify trends and patterns that can contribute to future improvements.

b) Availability Eliminating silos of road data. Some of the advantages are:

[b.1.] To introduce advanced monitoring for outlier (i.e., counterproductive error or data samples to improve the model robustness) detection across the system or a training framework (Jason et al., 2019).

[b.2.] To improve Machine-Learning Operations (MLOps) to deploy, learn, and optimise unexplored links between silos (Soh et al., 2020).

[b.3.] To enhance the systems interoperability through interfaces.

[b.3.] To enhance road safety measures by using previous road accident patterns across different regions

c) Completeness Creating different rules to automatically adapt these at different scales. Some of the features that are suggested to be explored further:

[c.1.] Data segmentation vs. aggregation.

[c.2.] Regular updates on data and processes.

[c.3.] Data storage.

d) Granularity Identify anomalies and detect changes. Some of the features suggested to be expanded are:

[d.1.] Identify the beginning of an anomaly.

[d.2.] To detect the most relevant variables in the occurrence of failure.

[d.3.] Define risk assessment metrics.

e) Relevance Improve the quality of the collected data. Some of the features that should be deployed further are:

[e.1.] To define a variety of metrics covering aspects of data collection.

[e.2.] To define measures to anticipate incorrect and incomplete data.

[e.3.] To define measures to discard duplicate data.

f) Reliability Improve the automation processes of the data quality. Some of the features suggested to be considered are:

[f.1.] Data collection processes.

[f.2.] Data synchronisation, to mitigate delays due to inaccurate data.

[f.3.] System and data quality inspections.

[f.4.] System and data risk assessments.

g) Standardisation Define international ML standards for road safety data. Some of the advantages are:

[g.1.] To establish standards on data processes (e.g., data entry, data filtering) and procedures (e.g., data audits).

[g.2.] To ensure data and process protection (e.g., ethics, confidentiality).

[g.3.] Creating data governance guidelines (Wiljer and Hakim, 2019)

[g.4.] Gaining more control over how to integrate the data.

[g.5.] To support data protection, security, and privacy issues.

h) Timeliness Monitoring and alerts to ensure standards and quality tests in near-real time. Some of the advantages:

[h.1.] Mitigate low quality data (e.g., misspelling errors).

[h.2.] Increase the security of data and processes.

[h.3.] Improve data transparency.

11. Conclusions and Future Work

11.1. Introduction

This study set out to systematically investigate the applicability of ML to improve pedestrian safety. Moreover, to discover new knowledge (e.g., trends, links) and assist among others, policymaking, road experts and road maintenance. This research applied an ML algorithm using a computational datamining tool to analyse pedestrian road-crash casualty data. This investigation has shown the degree of consistency of ML algorithm FCM in the form of a prototype as a development tool. The main research conclusions are presented under the following four headings:

11.2. Contributory Road Elements Identification

11.3. The Role of Actual Road Data

11.4. The FCM Prototype Feasibility

11.5. Transition Process into an operating FCM

11.2. Contributory Road Elements Identification

The conducted experiments confirmed that combining relevant road elements can lead to discover new insights. Moreover, the provision of these findings adds to the rapidly expanding field of conventional pedestrian safety analysis methods. In this study, the procedure to identify contributory road elements was carried out as follows [see Section 6.2.1]:

i. Pedestrian general attributes (i.e., gender, age group)

ii. Road environment (i.e., weather, speed limits, lighting)

iii. Road infrastructure (i.e., road type, pedestrian physical facilities)

To identify relevant elements more closely, it is of importance to highlight that in this study knowledge from quantitative and qualitative research was combined. As a result, it significantly strengthened the findings applicability.

11.3. The Role of Actual Road Data

The insights gained from gathering data on contributory road elements, emerged from the adopted data merging technique (i.e., pedestrian attributes, infrastructure). Disappointingly, to date fewer studies dealt with this topic. In terms of gathering requirements, even though a standardised procedure to analyse of pedestrian safety is not established yet, in this study, the analysis was performed by means of an individual set of data, interrelating the selected elements as follows:

i. A database storing the data (e.g., test, trial, field).

ii. A computational datamining tool to examine a limited road dataset for the desired analysis.

iii. A computational datamining tool handling the road data in a logical manner.

iv. A computational datamining tool visualising the road data in a graphical manner.

Similarly, to explore the road data more objectively (e.g., gather, filter, store), knowledge from the overall analysis, specific capabilities of the computational datamining tool were evaluated in more detail. Additionally, both the verification (e.g., FCM demonstrator) and the validation (i.e., FCM prototype) considerably improved the system analysis, respectively in terms of scalability for different regions (i.e., FCM demonstrator) and accuracy (e.g., ROC curve).

11.4. The FCM Prototype Feasibility

The FCM prototype has been deployed based on a full demonstration of the performance characteristics of an FCM applied to a few attributes, followed by an evaluation assessment including both verification and validation aspects. The analysis includes:

i. A database with road data for the applying of the FCM model to be used in pedestrian road data analysis through KNIME.

ii. The computational datamining tool KNIME, which is a derivative application from Java based on Eclipse, especially designed to perform modular data analysis.

iii. The applicability of adopting KDD to manage data and extract relevant knowledge consistently and efficiently manner.

iv. A rules system to assess relational features, based on complex statistical rules rather than pure arithmetic procedures using KNIME.

v. The use of data verification techniques to introduce a self-describing ability within the overall process.

Moreover, to demonstrate how such systems can be used to simulate the evaluation procedure performed during a trial or a field test, an elementary knowledge base has been automated with KNIME.

11.5. Transition Process into an Operating FCM

The transition process from the prototype into an operating FCM is referring to applying this approach at different times and in distinct geographical regions. To anticipate possible failures on future developments, standardisation provisions as the incorporation of the following items should be considered:

i. A procedural audit at each pre-defined process stage.

ii. A relational database supported by several facilities such as handling larger sets of files, availing a flexible statistical tool to model complex rules, and a geographical interface.

iii. An evaluation of the required memory (i.e., hardware and software) so that delays on the operability are reduced.

iv. A more user friendly interface (e.g., tool script writing and development) regarding the interface.

11.6. Overall Findings

Objective i. To identify features that affect pedestrian's road-crash casualties in relation to pedestrian general attributes. The research developed a datamining technique capable of analysing accident causation and extracting associated knowledge about any attribute included in the database. For the database considered (i.e., STATS19). The dataset used, considered all pedestrian accidents that occurred in 2017 in GB. This effected a spatial analysis of pedestrian safety across the London and GB in this instance. The use of data series for more than one year would be needed if a temporal analysis of data was sought. The following knowledge was extracted:

a. Pedestrian general attributes: human error was found a fundamental influential factor on pedestrian road-crash casualty causation. The results also indicate the validity of scrutinising information to improve pedestrian safety. A few remarks affecting pedestrian safety:

[a.1] On the gender and age group, the elderly, children and female were classified as the most risky. On the age group, data reported as non-recorded age were concluded as a critical disadvantage.

[a.2] To develop more realistic and scalable exploratory studies, restricting the number of participants per case study was concluded as a critical disadvantage.

[a.3] Limited literature on different levels of impairment, particularly regarding on cognitive disability was found.

b. Road environment: adverse road environmental conditions (e.g., weather, speed limits) were found as key influential factors on pedestrian road-crash casualties. A few key notes that can improve pedestrian safety:

[b.1] The implementation of physical facilities accounting the weather condition.

[b.2] Variable speed limits, and traffic calming treatments.

[b.3] Higher illuminance to reassure the pedestrians cognitive judgements.

c. Road infrastructure: the presence of pedestrian physical facilities (e.g., zebra crossings) is a fundamental factor to reduce pedestrian road-crash casualties.

[c.1] The island or median, yet, these share visibility limitations.

[c.2] Pedestrian safety models for the pelican, footbridge and private property were not found.

[c.3] The investigated pedestrian safety models revealed data discrepancies due to lack of standards.

Objective ii. To use STATS19 data and explore ML algorithms to describe and project safety trends for the pedestrians.

a. Pedestrian attributes: pedestrian general attributes assessed by gender and age group [a.1] Gender: the majority indicated (66.01%) chance of fatality for males. On the contrary, the minority indicated and a 33.99% chance for females.

[a.2] Age group: the majority indicated (37.07%) chance of fatality for elderly. By contrast, the minority indicated a 2.19% chance for young adolescents. However, a large number of non-recorded age rates (ca. 25.22%) were found.

b. Road environment: attributes assessed by the condition of the weather, speed limits and lighting)

[b.1] Weather: The majority indicated fine-no winds (FNW) with a 85.24% chance of fatality. On the contrary, the minority indicated Fine-high winds (FHW) with a 0.66% chance of fatality.

[b.2] Speed limits: the majority indicated a 69.12% chance of fatality for 30 mph. On the other hand, the minority indicated a 4.01% chance of fatality for 40 mph.

[b.3] Lighting: the majority indicated a 50.22% chance of fatality at daylight. On the contrary, the minority indicated a 1.1% chance of fatality at night with the lights unknown.

c. Road infrastructure: attributes assessed by road type and pedestrian physical facilities

[c.1] Road type: the majority indicated a 52.32% chance of fatality at no junction (NJ). On the contrary, the minority indicated a 0.68% chance of fatality at Multiple Junction (MJ).

[c.2] Pedestrian physical facilities: the majority accounted a 67.92% chance of fatality at no pedestrian physical facilities, and the minority indicated a 0.23 % chance of fatality at footbridge or subway.

Objective iii. To define a knowledge discovery approach for a safer pedestrian road environment. A knowledge discovery process was defined, and it consisted of the following components:

a. Pedestrian general attributes.

[a.1] Gender.

[a.2] Age Group.

b. Road environment.

[b.1] Weather.

[b.2] Speed Limit.

[b.3] Lighting.

c. Road infrastructure.

[c.1] Road type.

[c.2] Pedestrian physical facilities.

The process was developed using KNIME, a computational datamining tool to investigate patterns or undiscovered links. As part of the data analysis evaluation, it was felt that, maintainability, reliability, costs, certifiability were more related to evaluate a fully working system. Thus, only the following two characteristics were analysed: consistency, and sensitivity.

a. Consistency on the modular data flows.

[a.1] Pre-processing: converted road data from STATS19 into KNIME attributes.

[a.1] Classification: applied the FCM algorithm on actual data.

[a.2] Verification: reproduced the data in the same way in two different regions.

[a.3] Validation: applied the Spearman's rho and ROC curve to validate the algorithm in use.

b. Sensitivity response to actual data.

[b.1] New rules (i.e., 75%, 85%) to verify the FCM algorithm sensitivity response.

[b.2] Accuracy scoring rates were tested.

Objective iv. To test the applicability of the above methodology using data from the UK and from different regions. The applicability of the above methodology was tested.

Regarding the number of set criteria from adopted methodology, the potential of the FCM as a basis for investigating different regions was found. Moreover, because the knowledge from separable types of information (i.e., pedestrian general attributes, road environment, road infrastructure) can be simultaneously extracted.

a. The modular data flow demonstrator in different regions.

[a.1] Highest location of GB pedestrian fatalities: absent physical facilities.

[a.2] Highest region of London pedestrian fatalities: Enfield (i.e., 10)

b. Consistency of the modular data flow applicability on actual data.

[b.1] Pedestrian attributes: described the gender with the highest level of accuracy (i.e., female: 5.82%).

[b.2] Road environment: respectively indicated the lowest level of accuracy for speed limits (i.e., 30 mph: 42.69%) and the weather (i.e., fine no winds: 42.45%).

[b.3] Road infrastructure: described the lowest accuracy level at no Junction: 25.27%. However, interestingly it indicated a decreasing tendency on the chance of fatality risk in the case of absent pedestrian physical facilities.

c. Sensitivity of the modular data flow sensitivity response to actual data.

[c.1] Pedestrian attributes: by gender and by age group (especially for the elderly) the 75% set of rules represented the highest accuracy.

[c.2] Road environment: in overall, the indicative trends on weather, speed limits and lighting mostly fail. Exceptionally, the 85% set of rules for speed limits set at 70 mph mostly agreed with actual data.

[c.3] Road infrastructure: alike on the road environment, except for the 75% set of rules at roundabouts the calculation of the indicative trends by road type mostly failed. Nevertheless, particularly at roundabouts the findings incited promising trends.

Objective v. To review the literature related to road safety worldwide.

As part of this research study, the collected evidence was carried out worldwide. The findings indicate pedestrian physical facilities and other factors (e.g., human, speed limits), to play a crucial role on pedestrian safety. However, the literature findings do not explicitly indicate unknown patterns regarding pedestrian road-crash causal factors. To address issues alike, advancing cross-sectional studies to draw more conclusive trends, have been previously suggested (Urie et al., 2016; Useche et al., 2021). Therefore, examining this existing gap was of the central focus of this study.

Objective vi. To provide recommendations on road safety based on the data analysis and validate the findings with expert opinion.

To the knowledge of the author, the use of systematic methods can support reducing pedestrian road-crashes. Still, several questions remain to be answered. To help mitigating this challenge, in this study, the recommendations on pedestrian safety are provided by contrasting the findings from the literature review, the data analysis (i.e., temporal data), and the ML prototype (i.e., spatial data). In addition to that, to validate the findings, an online questionnaire to gather expert opinion was carried out.

11.7. Observations

The contribution of this study has been to confirm that systematic investigations using ML can improve pedestrian safety. However, the lack of data quality is considered the main weaknesses to deploy the prototype of this study. These results may partly be explained by human intervention affecting data interpretation, as well as the lack of standards to access relevant data (e.g., information on pedestrian physical facilities available is 2017, was neither available in 2018 nor in 2019). Nevertheless, this study has pioneered a thorough investigation on relational links between pedestrian attributes and road elements. As a result, it confirms the presence of pedestrian physical facilities as a primary factor in reducing pedestrian road-crash casualties. An emerging finding from this study is that interpreting pedestrian general attributes suggests being, if not an uncontrollable factor, highly challenging. It is difficult to explain this result, but it might be related to the complex nature of the subject, indicating reasons underpinning pedestrian road-crashes to remain unclear. Still, several factors could support understanding these reasons in more detail. For example, on the human factors, exploring the levels of cognition disabilities impacting on pedestrian safety. Moreover, another example, on data criteria, the selection of the most appropriate features (e.g., new rules) to examine the data is crucial. Nonetheless, the unsupervised ML spatial data analysis that was carried out using KNIME, relies on the existence of sufficient metadata to deploy a conceptual prototype. Furthermore, the deployed questionnaire to validate the prototype with expert opinion, enriched the findings of the prototype with new insights. Notwithstanding the relatively limited sample examined combined with the findings with expert opinion, this work offers valuable new knowledge to assist policymaking, decision-making and implementation procedures (e.g., road experts, road engineers, quality auditors, road data scientists). Subsequently, a larger number of road users can benefit from the recommendations provided in this study.

11.8. Recommendations

Based on the above conclusions the following recommendations may be made regarding the objectives of the thesis:

Recommendation i. Regarding pedestrian attributes, as human errors can affect pedestrian safety and especially that of the elderly, children and females, appropriate publicity, awareness campaigns, and educational training targeting specific groups of pedestrians (e.g., elderly, children, and females) should be developed. The study should be repeated for examining the influence of pedestrian infrastructure for pedestrians with distinct levels of cognitive disabilities. Turning into the road environment, since for adverse road environmental conditions (e.g., weather, speed limits) significantly affects pedestrian safety, it is recommended the implementation of facilities accounting for whether condition (e.g., side pavement ramps to access sidewalks), the introduction of variable speed limits (e.g., peak hours) alongside with traffic calming treatments. More research using controlled trials is recommended. Furthermore, to reassure the pedestrian judgements while commuting in a road environment, it may also be recommended to carry out a cross-national study to improve the levels of illuminance. With respect to road infrastructure, the presence of pedestrian physical facilities (e.g., zebra crossings) proved to be a fundamental factor that can improve pedestrian safety. Therefore, considerably more work will need to be done to develop a framework program and standardise the use of each of pedestrian physical facilities with the respective improvements included. For example, improving the limited pedestrian visibility on existing treatments (e.g., island or median). A natural progression of this work is to examine the effects of non-existing predictive models for pedestrian physical facilities (e.g., pelican, footbridge, private property). As regards data quality, a fruitful area for further work, is to mitigate data inconsistency. This means that, rather than adopting the conventional human intervention (e.g., a police officer capturing this same information manually), automating the data capturing process (e.g., near real-time data). It may also be recommended to create a system capable of integrating common data from different data sources (e.g., near real-time, hospitals, mortuaries) to assist interpreting data inaccuracies while enhancing the quality standards to access relevant data.

Recommendation ii. As regards the use of ML algorithms rules to describe and project pedestrian safety trends using STATS19 data, the prototype sensitivity analysis indicated the best accuracy rates to identify road-crash fatalities using different rules. By gender and by age group, the most accurate descriptor ratio was 75%, while by speed limits set at 70 mph the most accurate ratio was 85%. Thus, depending on the type of attributes under examination (e.g., pedestrian attributes, infrastructure), instead of adopting ML algorithms using traditional rules to train the data (i.e., descriptor: 80%), for future work, it may be recommended to examine different rules (e.g., Gender- descriptor: 75%, predictor: 15%; Speed limits - descriptor: 85%, predictor: 15%). Moreover, for different regions (e.g., Ireland, and New Zealand) larger randomised controlled trials could provide more definitive evidence.

Recommendation iii. Regarding the knowledge discovery process, a greater focus on promoting a decision-making engagement activity based on the findings examination and integration (i.e., literature review, prototype, and expert opinion) could produce interesting findings that account more for policy-making initiatives.

Recommendation iv. As regards testing the applicability of the methodology using data from the UK and from different regions, data quality was noticed to be vital. Therefore, it may be recommended improve data quality and consistency by introducing a standardised homogeneous framework that accounts data from the GB and from different regions.

Recommendation v. Regarding the literature review on pedestrian safety worldwide, it may be recommended to deploy a more a precise mechanism to implement relevant road treatments, for different levels of impaired roads users (e.g., variable speed limits at specific junctions, new concepts of pedestrian facilities, new policy for pedestrians with dementia). Moreover, these findings provide the following insights for future research: the impact of pedestrians by gender, age group, and with distinct levels and types of cognitive abilities in a road environment.

Recommendation vi. On the data analysis, the validation with the expert opinion was successfully achieved. This since, the focus group participants expressed their interest and willingness by engaging and contributing to the questionnaire. On one hand, the prototype proved its applicability. However, as on some occasions the prototype accuracy measures may be under performing, the extracted information may lead to uncertainty. Further research should focus on determining the links between accuracy measures and uncertainty. Taken together, the findings from quantitative research (i.e., prototype) and qualitative research (i.e., questionnaire) methods demonstrated to be complementary. Therefore, to improve pedestrian safety the adoption of a mixed method of research (i.e., quantitative vs. qualitative), is highly recommended [see Chapters 2, 8, and 10]. Furthermore, it is suggested that before carrying out with further research on pedestrians with cognition disabilities is introduced, a study like this one should be carried out on assessing, and re-evaluating existing metadata, locally, regionally, and globally. This since, the prototype demonstrated its applicability to different regions. Subsequently, it can become quite useful to effectively expand future discoveries to a wider magnitude.

11.9. Future Work

In this study, an ML proof-of-concept was deployed to assess pedestrian safety data. For future work it may be recommended to incorporate additional contributory factors (e.g., cognitive disabled road users, cyclists' general attributes, carriageway hazards, road surface condition) influencing pedestrian road crashes on the examination. Additionally, make use of real-time data to evaluate the suitability of the existing database STATS19 to collect data from road-crashes. Turning into the ML analysis, the pedestrian safety knowledge discovery process faced some limitations due to among others, human error, lack of data quality and consistency. Therefore, further investigation on pedestrian road data using predictive modelling is strongly recommended. Several measures that future research might explore are presented as follows:

1. Advancing a multi-layer modelling method so that a wider number of road elements can be combined into a single analysis. For example, additional road users (e.g., drivers, cyclists, or motorcyclists), or road physical facilities (e.g., hazardous roads)

2. Although WEKA was initially experimented as a testing tool, in this study only one modular ML application, KNIME, was further explored. Therefore, to investigate road data, testing and using of other ML application systems more comprehensively in the market is highly recommended.

3. This study gave focus to a systematic investigation of the ML applicability to improve pedestrian safety in an urban road environment. Nevertheless, con-

tinued efforts are needed. It was also felt that it would be useful to develop ML algorithms that include examining the impact of the behaviour of other road users (e.g., drivers, cyclists, moppers, or motorcyclists) located in road facilities where predictive models have not been explored to date (e.g., pelican, footbridge, and private property). Additionally, developing a method that, based on the road environment (e.g., rural, and urban), can help assess, scale, and explore road features, is highly recommended. Moreover, expand the effects of the predictor sub-mode (e.g., explore the unknown) to the overall analysis.

4. The deployed questionnaire in this study was designated for experts in the field of engineering. Future initiatives assigned to a wider community with focus groups from different fields of expertise (e.g., health, data) and to focus groups with different types of road users (e.g., male pedestrians, pedestrians with a physical level of impairment) are suggested.

Bibliography

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, *76*, 243–297.

Addakiri, K., Khallouki, H., & Bahaj, M. (2019). Intelligent urban transport decision analysis system based on mining in big data analytics and data warehouse. *International Conference on Artificial Intelligence and Symbolic Computation*, 179–184.

Adminaite, D., Allsop, R., & Jost, G. (2015). Ranking eu progress on road safety: 9th road safety performance index report.

Agran, P., Winn, D., & Dunkle, D. (1989). Injuries among 4-to 9-year-old restrained motor vehicle occupants by seat location and crash impact site. *American Journal of Diseases of Children*, *143*(11), 1317–1321.

Al Bargi, W. A., & Daniel, B. D. (2020). Modelling pedestrians' utilization of crossing facilities along urban streets. *Case studies on transport policy*, *8*(2), 593–598.

Allen, D. E., Singh, A. K., & Powell, R. J. (2011). Extreme market risk-an extreme value theory approach.

Ampofo-Boateng, K., & Thomson, J. A. (1991). Children's perception of safety and danger on the road. *British Journal of Psychology*, *82*(4), 487–505.

Anciaes, P. R., & Jones, P. (2018). Estimating preferences for different types of pedestrian crossing facilities. *Transportation research part F: traffic psychology and behaviour*, *52*, 222–237.

Anderson, R., Taylor, S., Taylor, T., & Virues-Ortega, J. (2022). Thematic and textual analysis methods for developing social validity questionnaires in applied behavior analysis. *Behavioral Interventions*, *37*(3), 732–753.

Ang, A., Christensen, P., & Vieira, R. (2020). Should congested cities reduce their speed limits? evidence from são paulo, brazil. *Journal of Public Economics*, *184*, 104155.

Assi, K., Rahman, S. M., Mansoor, U., & Ratrout, N. (2020). Predicting crash injury severity with machine learning algorithm synergized with clustering technique: A promising protocol. *International journal of environmental research and public health*, *17*(15), 5497.

Austroads. (2019). Austroads annual report 2018-19.

Ayers, J. W., Leas, E. C., Dredze, M., Allem, J.-P., Grabowski, J. G., & Hill, L. (2016). Pokémon go—a new distraction for drivers and pedestrians. *JAMA internal medicine*, *176*(12), 1865–1866.

Aziz, H. A., Ukkusuri, S. V., & Hasan, S. (2013). Exploring the determinants of pedestrian–vehicle crash severity in new york city. *Accident Analysis & Prevention*, *50*, 1298–1309.

Ballesteros, M. F., Dischinger, P. C., & Langenberg, P. (2004). Pedestrian injuries and vehicle type in maryland, 1995–1999. *Accident Analysis & Prevention*, *36*(1), 73–81.

Bastien, C., Neal-Sturgess, C., Christensen, J., & Wen, L. (2020). A method to calculate the ais trauma score from a finite element model. *Journal of Mechanics in Medicine and Biology*, *20*(06), 2050034.

Bauer, R., Machata, K., Brandstaetter, C., Yannis, G., Laiou, A., & Folla, K. (2016). Road traffic accidents in european urban areas. *Proceedings of the 1st European Road Infrastructure Congress, Leeds, UK*, 18–20.

Bentley, T. A., & Haslam, R. (2001). Identification of risk factors and countermeasures for slip, trip and fall accidents during the delivery of mail. *Applied ergonomics*, *32*(2), 127–134.

Beshah, T., & Hill, S. (2010). Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in ethiopia. *2010 AAAI Spring symposium series*.

Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.

Bhat, C. R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation science*, *31*(1), 34–48.

Bichicchi, A., Mazzotta, F., Lantieri, C., Vignali, V., Simone, A., Dondi, G., & Costa, M. (2017). The influence of pedestrian crossings features on driving behavior and road safety. In *Transport infrastructure and systems* (pp. 741–746). CRC Press.

Boarnet, M. G., Day, K., Anderson, C., McMillan, T., & Alfonzo, M. (2005). California's safe routes to school program: Impacts on walking, bicycling, and pedestrian safety. *Journal of the American Planning Association*, *71*(3), 301–317.

Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20—79 years: Reference values and determinants. *Age and ageing*, *26*(1), 15–19.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Bullough, J. D., Donnell, E. T., & Rea, M. S. (2013). To illuminate or not to illuminate: Roadway lighting as it affects traffic safety at intersections. *Accident Analysis & Prevention*, *53*, 65–77.

Bungum, T. J., Day, C., & Henry, L. J. (2005). The association of distraction and caution displayed by pedestrians at a lighted crosswalk. *Journal of community health*, *30*(4), 269–279.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, *2*(2), 121–167.

Buscema, M., Breda, M., & Lodwick, W. (2013). Training with input selection and testing (twist) algorithm: A significant advance in pattern recognition performance of machine learning.

Butsenko, E. (2022). How data mining can improve road safety in cities. *Social Sciences*, *11*(3), 130.

Cadar, R. D., Boitor, M. R., & Dumitrescu, M. (2017). Effects of traffic volumes on accidents: The case of romania's national roads. *Geogr. Tech*, *12*(2), 20–29.

Cahill, M. (2010). *Transport, environment and society* (Vol. 11). McGraw-Hill Education (UK).

Cameron, M. H., & Elvik, R. (2010). Nilsson's power model connecting speed and road trauma: Applicability by road type and alternative models for urban roads. *Accident Analysis & Prevention*, *42*(6), 1908–1915.

Casado-Sanz, N., Guirao, B., & Gálvez-Pérez, D. (2019). Population ageing and rural road accidents: Analysis of accident severity in traffic crashes with older pedestrians on spanish crosstown roads. *Research in Transportation Business & Management*, *30*, 100377.

Castillo-Manzano, J. I., Castro-Nuño, M., Fageda, X., & Lopez-Valpuesta, L. (2017). An assessment of the effects of alcohol consumption and prevention policies on traffic fatality rates in the enlarged eu. time for zero alcohol tolerance? *Transportation research part F: traffic psychology and behaviour*, *50*, 38–49.

Chandra, S., & Bharti, A. K. (2013). Speed distribution curves for pedestrians during walking and crossing. *Procedia-Social and Behavioral Sciences*, *104*, 660–667.

Chen, Y., Meng, H., & Wang, Z. (2009). *Safety improvement practice for vulnerable road users in beijing junctions* (tech. rep.).

Chong, M., Abraham, A., & Paprzycki, M. (2004). Traffic accident data mining using machine learning paradigms. *Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary*, 415–420.

Clayton, A. B., & Colgan, M. (2001). Alcohol and pedestrians-a final report to road safety division. *Road Safety Research Report 20.*

Cohen, A., Bar-Gera, H., Parmet, Y., & Ronen, A. (2013). Guardrail influence on pedestrian crossing behavior at roundabouts. *Accident Analysis & Prevention*, *59*, 452–458.

Cooper, M. (2022). Longer males determined with positive skew and kurtosis in centrobolus (diplopoda: Spirobolida: Pachybolidae). *New Visions in Biological Science Vol*, *8*, 102–106.

Cordellieri, P., Sdoia, S., Ferlazzo, F., Sgalla, R., & Giannini, A. M. (2019). Driving attitudes, behaviours, risk perception and risk concern among young student car-drivers, motorcyclists and pedestrians in various eu countries. *Transportation research part F: traffic psychology and behaviour*, *65*, 56–67.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons.

Costa, M., Bonetti, L., Vignali, V., Lantieri, C., & Simone, A. (2018). The role of peripheral vision in vertical road sign identification and discrimination. *Ergonomics*, *61*(12), 1619–1634.

Cottrell, M., & Rousset, P. (1997). The kohonen algorithm: A powerful tool for analysing and representing multidimensional quantitative and qualitative data. *International Work-Conference on Artificial Neural Networks*, 861–871.

DaSilva, M., Smith, J., & Najm, W. (2003). *Analysis of pedestrian crashes* (tech. rep.).

De Ona, J., López, G., Mujalli, R., & Calvo, F. J. (2013). Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. *Accident Analysis & Prevention*, *51*, 1–10.

De Winter, J. C., Gosling, S. D., & Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, *21*(3), 273.

Deb, S., Strawderman, L., Carruth, D. W., DuBien, J., Smith, B., & Garrison, T. M. (2017). Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. *Transportation research part C: emerging technologies*, *84*, 178–195.

Dellinger, A. M., for Disease Control, C., Prevention, et al. (2002). Barriers to children walking and biking to school–united states, 1999. *MMWR: Morbidity and mortality weekly report*, *51*(32), 701–704.

DeLucia, P. R., Bleckley, M. K., Meyer, L. E., & Bush, J. M. (2003). Judgments about collision in younger and older drivers. *Transportation research part F: traffic psychology and behaviour*, *6*(1), 63–80.

Demetriades, D., Murray, J., Charalambides, K., Alo, K., Velmahos, G., Rhee, P., & Chan, L. (2004). Trauma fatalities: Time and location of hospital deaths. *Journal of the American College of Surgeons*, *198*(1), 20–26.

DfT. (2011). Department for transport: Instructions for the completion of road accident reports from non-crash sources.

DfT. (2013). Department for transport: Road accidents and safety statistics database. great britain.

DfT. (2016). Department for transport: Reported road casualties great britain: 2016 annual report.

DfT. (2017). Department for transport: Reported road casualties great britain: 2017 annual report.

DfT. (2018). Department for transport: Reported road casualties in great britain:2018 annual report.

DfT. (2019). Department for transport: Reported road casualties in great britain: 2019 annual report.

DfT. (2020). Department for transport:reported road casualties in great britain: 2020 annual report: Provisional results.

Drosu, A., & Cofaru, C. (2016). Estimating the costs caused by road traffic accidents in romania. *International Congress of Automotive and Transport Engineering*, 889–898.

Dultz, L. A., & Frangos, S. G. (2013). The impact of alcohol in pedestrian trauma. *Trauma*, *15*(1), 64–75.

Earl, R., Falkmer, T., Girdler, S., Morris, S. L., & Falkmer, M. (2018). Viewpoints of pedestrians with and without cognitive impairment on shared zones and zebra crossings. *PLoS one*, *13*(9), e0203765.

Earth, G. (2021a). Pedestrian fatalities in gb by road type.

Earth, G. (2021b). Pedestrian fatalities in london by local authority. *Google. (n.d.)*

Ehsaei, A., & Evdorides, H. (2011). Temporal variation of road accident data caused by road infrastructure. *3rd International Conference on Road Safety and SimulationPurdue UniversityTransportation Research Board.*

Elaziz, M. A., Hosny, K. M., Salah, A., Darwish, M. M., Lu, S., & Sahlol, A. T. (2020). New machine learning method for image-based diagnosis of covid-19. *Plos one*, *15*(6), e0235187.

Elias, W., Toledo, T., & Shiftan, Y. (2010). The effect of daily-activity patterns on crash involvement. *Accident Analysis & Prevention*, *42*(6), 1682–1688.

Elliott, S., Woolacott, H., & Braithwaite, R. (2009). The prevalence of drugs and alcohol found in road traffic fatalities: A comparative study of victims. *Science & Justice*, *49*(1), 19–23.

Elton, D. C., Boukouvalas, Z., Butrico, M. S., Fuge, M. D., & Chung, P. W. (2018). Applying machine learning techniques to predict the properties of energetic materials. *Scientific reports*, *8*(1), 1–12.

Eluru, N., Bagheri, M., Miranda-Moreno, L. F., & Fu, L. (2012). A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. *Accident Analysis & Prevention*, *47*, 119–127.

Elvik, R., Vaa, T., Hoye, A., & Sorensen, M. (2009). *The handbook of road safety measures*. Emerald Group Publishing.

Embrey, D. (2002). Using influence diagrams to analyse and predict failures in safety critical systems. *23rd ESReDA Seminar-Decision Analysis: Methodology and Applications for Safety of Transportation and Process Industries*.

Feliciani, C., Crociani, L., Gorrini, A., Vizzari, G., Bandini, S., & Nishinari, K. (2017). A simulation model for non-signalized pedestrian crosswalks based on evidence from on field observation. *Intelligenza Artificiale*, *11*(2), 117–138.

Feliciani, C., Crociani, L., Gorrini, A., Vizzari, G., Nishinari, K., & Bandini, S. (2017). Assessment of pedestrian fatality risk at unsignalized crosswalks by means of simulation. *International Conference on Traffic and Granular Flow*, 423–431.

Feliciani, C., Gorrini, A., Crociani, L., Vizzari, G., Nishinari, K., & Bandini, S. (2020). Calibration and validation of a simulation model for predicting pedestrian fatalities at unsignalized crosswalks by means of statistical traffic data. *Journal of traffic and transportation engineering (English edition)*, *7*(1), 1–18.

Fiorentini, N., & Losa, M. (2020). Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructures*, *5*(7), 61.

Fischer, E., & Guzel, G. T. (2023). The case for qualitative research. *Journal of Consumer Psychology*, *33*(1), 259–272.

Forum., I. T. (2020). Road safety annual report 2020.

Fotios, S., & Castleton, H. (2016). Specifying enough light to feel reassured on pedestrian footpaths. *Leukos*, *12*(4), 235–243.

Fotios, S., & Gibbons, R. (2018). Road lighting research for drivers and pedestrians: The basis of luminance and illuminance recommendations. *Lighting Research & Technology*, *50*(1), 154–186.

Gao, X., Zhou, Y., Amir, M. I. H., Rosyidah, F. A., & Lee, G. M. (2017). A hybrid genetic algorithm for multi-emergency medical service center location-allocation problem in disaster response. *International Journal of Industrial Engineering*, *24*(6).

Gariazzo, C., Stafoggia, M., Bruzzone, S., Pelliccioni, A., & Forastiere, F. (2018). Association between mobile phone traffic volume and road crash fatalities: A population-based case-crossover study. *Accident Analysis & Prevention*, *115*, 25–33.

Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, *4*(4).

Giles-Corti, B., Wood, G., Pikora, T., Learnihan, V., Bulsara, M., Van Niel, K., Timperio, A., McCormack, G., & Villanueva, K. (2011). School site and the potential to walk to school: The impact of street connectivity and traffic exposure in school neighborhoods. *Health & place*, *17*(2), 545–550.

Gitelman, V., Balasha, D., Carmel, R., Hendel, L., & Pesahov, F. (2012). Characterization of pedestrian accidents and an examination of infrastructure measures to improve pedestrian safety in israel. *Accident Analysis & Prevention*, *44*(1), 63–73.

Giummarra, M. J., Beck, B., & Gabbe, B. J. (2021). Classification of road traffic injury collision characteristics using text mining analysis: Implications for road injury prevention. *PloS one*, *16*(1), e0245636.

Godavarthy, R. P., & Russell, E. R. (2016). Study of pedestrian hybrid beacon's effectiveness for motorists at midblock pedestrian crossings. *Journal of Traffic and Transportation Engineering (English Edition)*, *3*(6), 531–539.

González-González, E., Cordera, R., Alonso, B., & Nogués, S. (2021). Planning for sustainable urban mobility: Demand estimation of on-street vertical walking facilities. *International Journal of Sustainable Transportation*, 1–23.

Gorrini, A., Crociani, L., Vizzari, G., & Bandini, S. (2018). Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation research part F: traffic psychology and behaviour*, *59*, 269–285.

Graham, D., Glaister, S., & Anderson, R. (2005). The effects of area deprivation on the incidence of child and adult pedestrian casualties in england. *Accident Analysis & Prevention*, *37*(1), 125–135.

Graham, D. J., & Glaister, S. (2003). Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix. *Urban Studies*, *40*(8), 1591–1607.

Granà, A., Giuffrè, T., & Guerrieri, M. (2010). Exploring effects of area-wide traffic calming measures on urban road sustainable safety. *Journal of Sustainable Development*, *3*(4), 38.

Griselda, L., Joaquín, A., et al. (2012). Using decision trees to extract decision rules from police reports on road accidents. *Procedia-social and behavioral sciences*, *53*, 106–114.

Gross, F., Lyon, C., Persaud, B., & Srinivasan, R. (2013). Safety effectiveness of converting signalized intersections to roundabouts. *Accident Analysis & Prevention*, *50*, 234–241.

Gu, J., Jiao, L., Yang, S., & Zhao, J. (2017). Sparse learning based fuzzy c-means clustering. *Knowledge-based systems*, *119*, 113–125.

Guth, D. A., Long, R. G., Wall Emerson, R. S., Ponchillia, P. E., & Ashmead, D. H. (2013). Blind and sighted pedestrians' road-crossing judgments at a single-lane roundabout. *Human factors*, *55*(3), 632–642.

Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review. *Journal of traffic and transportation engineering (English edition)*.

Haddon, W. (1970). *Why the issue is loss reduction rather than only crash prevention* (tech. rep.). SAE Technical Paper.

Häkkänen, L., Ilgın, H. E., & Karjalainen, M. (2022). The current state of the finnish cottage phenomenon: Perspectives of experts. *Buildings*, *12*(3), 260.

Hamann, C., Peek-Asa, C., & Rus, D. (2015). Epidemiology of pedestrian–mvcs by road type in cluj, romania. *Injury prevention*, *21*(2), 84–90.

Han, J., Kamber, M., & Pei, J. (2012). Outlier detection. *Data mining: concepts and techniques*, 543–584.

Hariyono, J., & Jo, K.-H. (2017). Detection of pedestrian crossing road: A study on pedestrian pose recognition. *Neurocomputing*, *234*, 144–153.

Hatfield, J., & Murphy, S. (2007). The effects of mobile phone use on pedestrian crossing behaviour at signalised and unsignalised intersections. *Accident analysis & prevention*, *39*(1), 197–205.

Helmers, G., Henriksson, P., & Hakamies-Blomqvist, L. (2004). *Trafikmiljö för äldre bilförare: Analys och rekommendationer utifrån en litteraturstudie*. Statens väg-och transportforskningsinstitut., VTI rapport 493.

Hezaveh, A. M., & Cherry, C. R. (2018). Walking under the influence of the alcohol: A case study of pedestrian crashes in tennessee. *Accident Analysis & Prevention*, *121*, 64–70.

Hiebl, J., & Frei, C. (2016). Daily temperature grids for austria since 1961—concept, creation and applicability. *Theoretical and applied climatology*, *124*(1), 161–178.

Holland, C., & Hill, R. (2007). The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations. *Accident Analysis & Prevention*, *39*(2), 224–237.

Hou, M., Cheng, J., Xiao, F., & Wang, C. (2021). Distracted behavior of pedestrians while crossing street: A case study in china. *International journal of environmental research and public health*, *18*(1), 353.

Hoy, M., Matveev, A. S., & Savkin, A. V. (2015). Algorithms for collision-free navigation of mobile robots in complex cluttered environments: A survey. *Robotica*, *33*(3), 463–497.

Huang, H., Zegeer, C., & Nassi, R. (2000). Innovative treatments at unsignalized pedestrian crossing locations. *ITE journal, online tilgjengelig på http://www. ite. org/traffic/documents/AB00H5102. pdf.*

Hydén, C., & Várhelyi, A. (2000). The effects on safety, time consumption and environment of large scale use of roundabouts in an urban area: A case study. *Accident Analysis & Prevention*, *32*(1), 11–23.

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. *Accident Analysis & Prevention*, *130*, 84–90.

Infante, P., Jacinto, G., Afonso, A., Rego, L., Nogueira, P., Silva, M., Nogueira, V., Saias, J., Quaresma, P., Santos, D., et al. (2023). Factors that influence the type of road traffic accidents: A case study in a district of portugal. *Sustainability*, *15*(3), 2352.

Injured and killed in road traffic accidents by type of accident, casualty, transport unit, position, type of person and influenced by alcohol. (2020). *StatBank Denmark*.

Jaber, A. N., & Rehman, S. U. (2020). Fcm–svm based intrusion detection system for cloud computing environment. *Cluster Computing*, 1–11.

Jagnoor, J., & Peden, M. (2020). Injuries as global health risk factor: Causes, burden and prevention. *Handbook of Global Health*, 1–31.

Jain, A., Gupta, A., & Rastogi, R. (2014). Pedestrian crossing behaviour analysis at intersections. *International Journal for Traffic and Transport Engineering*, *4*(1), 103–116.

James, D. E., Schraw, G., & Kuch, F. (2015). Using the sampling margin of error to assess the interpretative validity of student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *40*(8), 1123–1141.

Javaid, A., Dardari, Z. A., Mitchell, J. D., Whelton, S. P., Dzaye, O., Lima, J. A., Lloyd-Jones, D. M., Budoff, M., Nasir, K., Berman, D. S., et al. (2022). Distribution of coronary artery calcium by age, sex, and race among patients 30-45 years old. *Journal of the American College of Cardiology*, *79*(19), 1873–1886.

Jomnonkwao, S., Uttra, S., & Ratanavaraha, V. (2021). Analysis of a driving behavior measurement model using a modified driver behavior questionnaire encompassing texting, social media use, and drug and alcohol consumption. *Transportation research interdisciplinary perspectives*, *9*, 100302.

Kangas, M., Heikinheimo, M., & Hippi, M. (2015). Roadsurf: A modelling system for predicting road weather and road surface conditions. *Meteorological applications*, *22*(3), 544–553.

Kann, A., Pistotnik, G., & Bica, B. (2012). Inca-ce: A central european initiative in nowcasting severe weather and its applications. *Advances in Science and Research*, *8*(1), 67–75.

Kaparias, I., Bell, M. G., Miri, A., Chan, C., & Mount, B. (2012). Analysing the perceptions of pedestrians and drivers to shared space. *Transportation research part F: traffic psychology and behaviour*, *15*(3), 297–310.

Kaplan, S., & Prato, C. G. (2012). Risk factors associated with bus accident severity in the united states: A generalized ordered logit model. *Journal of safety research*, *43*(3), 171–180.

Keegan, O., & O'Mahony, M. (2003). Modifying pedestrian behaviour. *Transportation Research Part A: Policy and Practice*, *37*(10), 889–901.

Khan, K., Zaidi, S. B., & Ali, A. (2020). Evaluating the nature of distractive driving factors towards road traffic accident. *Civil Engineering Journal*, *6*(8), 1555–1580.

Kim, J.-K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2011). *Mixed logit model approach to investigate effects of age on driver injury severity in single-vehicle accidents* (tech. rep.).

Kim, K., Brunner, I. M., & Yamashita, E. (2008). Modeling fault among accident—involved pedestrians and motorists in hawaii. *Accident Analysis & Prevention*, *40*(6), 2043–2049.

Klenk, J., Becker, C., Lieken, F., Nicolai, S., Maetzler, W., Alt, W., Zijlstra, W., Hausdorff, J. M., Van Lummel, R., Chiari, L., et al. (2011). Comparison of acceleration signals of simulated and real-world backward falls. *Medical engineering & physics*, *33*(3), 368–373.

KNIME. (2021). Konstanz information miner. *KNIME (n.d.)*

Koh, P., Wong, Y., & Chandrasekar, P. (2014). Safety evaluation of pedestrian behaviour and violations at signalised pedestrian crossings. *Safety science*, *70*, 143–152.

Kuhn, M., & Johnson, K. (2013). Linear regression and its cousins. In *Applied predictive modeling* (pp. 101–139). Springer.

Kwon, O. H., Rhee, W., & Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis & Prevention*, *75*, 1–15.

Lanera, C., Berchialla, P., Sharma, A., Minto, C., Gregori, D., & Baldi, I. (2019). Screening pubmed abstracts: Is class imbalance always a challenge to machine learning? *Systematic reviews*, *8*(1), 1–9.

Larson, S., Mahendran, A., Lee, A., Kummerfeld, J. K., Hill, P., Laurenzano, M. A., Hauswald, J., Tang, L., & Mars, J. (2019). Outlier detection for improved data quality and diversity in dialog systems. *arXiv preprint arXiv:1904.03122*.

Lasota, D., Goniewicz, M., Kosson, D., Ochal, A., Krajewski, P., Tarka, S., Goniewicz, K., & Mirowska-Guzel, D. (2019). The effect of ethyl alcohol on the severity of injuries in fatal pedestrian victims of traffic crashes. *PLoS one*, *14*(9), e0221749.

Lee, C., & Abdel-Aty, M. (2005). Comprehensive analysis of vehicle–pedestrian crashes at intersections in florida. *Accident Analysis & Prevention*, *37*(4), 775–786.

Lee, H., & Kim, S.-N. (2019). Shared space and pedestrian safety: Empirical evidence from pedestrian priority street projects in seoul, korea. *Sustainability*, *11*(17), 4645.

Li, C., Liu, S., & Cen, X. (2021). Safety and efficiency impact of pedestrian–vehicle conflicts at non signalized midblock crosswalks based on fuzzy cellular automata. *Physica A: Statistical Mechanics and its Applications*, *572*, 125871.

Li, F., Chen, C.-H., Zheng, P., Feng, S., Xu, G., & Khoo, L. P. (2020). An explorative context-aware machine learning approach to reducing human fatigue risk of traffic control operators. *Safety science*, *125*, 104655.

Li, H., Raeside, R., Chen, T., & McQuaid, R. W. (2012). Population ageing, gender and the transportation system. *Research in transportation economics*, *34*(1), 39–47.

Li, Q., Wang, Z., Yang, J., & Wang, J. (2005). Pedestrian delay estimation at signalized intersections in developing cities. *Transportation Research Part A: Policy and Practice*, *39*(1), 61–73.

Lin, P.-S., Guo, R., Bialkowska-Jelinska, E., Kourtellis, A., & Zhang, Y. (2019). Development of countermeasures to effectively improve pedestrian safety in low-income areas. *Journal of traffic and transportation engineering (English edition)*, *6*(2), 162–174.

Lluri, L., & Golgota, A. (2019). Analysis of road safety on the secondary road network in albania. *Studime sociale*, *13*(3).

Lyon, C., & Persaud, B. (2002). Pedestrian collision prediction models for urban intersections. *Transportation research record*, *1818*(1), 102–107.

Ma, M., Yan, X., Huang, H., & Abdel-Aty, M. (2010). Safety of public transportation occupational drivers: Risk perception, attitudes, and driving behavior. *Transportation Research Record*, *2145*(1), 72–79.

Malin, F., & Luoma, J. (2020). Effects of speed display signs on driving speed at pedestrian crossings on collector streets. *Transportation research part F: traffic psychology and behaviour*, *74*, 433–438.

Malkhamah, S., Tight, M., & Montgomery, F. (2005). The development of an automatic method of safety monitoring at pelican crossings. *Accident Analysis & Prevention*, *37*(5), 938–946.

Marino, F., Leccese, F., & Pizzuti, S. (2017). Adaptive street lighting predictive control. *Energy Procedia*, *111*, 790–799.

Marisamynathan, S., & Vedagiri, P. (2019). Pedestrian perception-based level-of-service model at signalized intersection crosswalks. *Journal of modern transportation*, *27*(4), 266–281.

Martin, J.-L., & Wu, D. (2018). Pedestrian fatality and impact speed squared: Cloglog modeling from french national data. *Traffic injury prevention*, *19*(1), 94–101.

Martín, L., Baena, L., Garach, L., López, G., & De Oña, J. (2014). Using data mining techniques to road safety improvement in spanish roads. *Procedia-social and behavioral sciences*, *160*, 607–614.

Maryam, H., Panayiotou, T., & Ellinas, G. (2023). Uncertainty quantification and consideration in ml-aided traffic-driven service provisioning. *Computer Communications*, *202*, 13–22.

Massaro, A., & Galiano, A. (2020). Re-engineering process in a food factory: An overview of technologies and approaches for the design of pasta production processes. *Production & Manufacturing Research*, *8*(1), 80–100.

Mayora, J. M. P., & Piña, R. J. (2009). An assessment of the skid resistance effect on traffic safety under wet-pavement conditions. *Accident Analysis & Prevention*, *41*(4), 881–886.

Mayou, R., & Bryant, B. (2003). Consequences of road traffic accidents for different types of road user. *Injury*, *34*(3), 197–202.

Mendoza, Q. A., Pordesimo, L., Neilsen, M., Armstrong, P., Campbell, J., & Mendoza, P. T. (2023). Application of machine learning for insect monitoring in grain facilities. *AI*, *4*(1), 348–360.

Meyer, D., Muir, S., Silva, S. S. M., Slikboer, R., McIntyre, A., Imberger, K., & Pyta, V. (2021). Modelling the relationship of driver license and offense history with fatal and serious injury (fsi) crash involvement. *Journal of safety research*.

Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. *Accident Analysis & Prevention*, *40*(1), 260–266.

Mirhashemi, A., Amirifar, S., Kashani, A. T., & Zou, X. (2022). Macro-level literature analysis on pedestrian safety: Bibliometric overview, conceptual frames, and trends. *Accident Analysis & Prevention*, *174*, 106720.

Mocan, R., & Dioşan, L. (2016). Multiclass classification based on clustering approaches for obstacle recognition in traffic scenes. *2016 IEEE 12th Interna-*

tional Conference on Intelligent Computer Communication and Processing (ICCP), 257–261.

Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., & Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in new york, us and montreal, canada. *Safety science*, *54*, 27–37.

Mohan, D., Tiwari, G., & Mukherjee, S. (2016). Urban traffic safety assessment: A case study of six indian cities. *IATSS research*, *39*(2), 95–101.

Mohanty, M., & Samal, S. R. (2019). „role of young drivers in road crashes: A case study in india". *European Transport\Trasporti Europei*, (74).

Montufar, J., Arango, J., Porter, M., & Nakagawa, S. (2007). Pedestrians' normal walking speed and speed when crossing a street. *Transportation research record*, *2002*(1), 90–97.

Muchanga-Hvelplund, L. C. (2023). Questionnaire: The applicability of ÄI to improve pedestrian safety.

Mura, P., Chatelain, C., Dumestre, V., Gaulier, J., Ghysel, M., Lacroix, C., Kergueris, M., Lhermitte, M., Moulsma, M., Pepin, G., et al. (2006). Use of drugs of abuse in less than 30-year-old drivers killed in a road crash in france: A spectacular increase for cannabis, cocaine and amphetamines. *Forensic science international*, *160*(2-3), 168–172.

Mwakalonge, J., Siuhi, S., & White, J. (2015). Distracted walking: Examining the extent to pedestrian safety problems. *Journal of traffic and transportation engineering (English edition)*, *2*(5), 327–337.

Naik, B., Tung, L.-W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury severity. *Journal of safety research*, *58*, 57–65.

Nashad, T., Yasmin, S., Eluru, N., Lee, J., & Abdel-Aty, M. A. (2016). Joint modeling of pedestrian and bicycle crashes: Copula-based approach. *Transportation research record*, *2601*(1), 119–127.

Nemire, K., Johnson, D. A., & Vidal, K. (2016). The science behind codes and standards for safe walkways: Changes in level, stairways, stair handrails and slip resistance. *Applied ergonomics*, *52*, 309–316.

Nhac-Vu, H.-T., Hours, M., Chossegros, L., Charnay, P., Tardy, H., Martin, J.-L., Mazaux, J.-M., & Laumon, B. (2014). Prognosis of outcome in adult survivors of road accidents in france: One-year follow-up in the esparr cohort. *Traffic injury prevention*, *15*(2), 138–147.

Niebuhr, T., Junge, M., & Rosen, E. (2016). Pedestrian injury risk and the effect of age. *Accident Analysis & Prevention*, *86*, 121–128.

Nilsson, G. (2004). *Traffic safety dimensions and the power model to describe the effect of speed on safety* [Doctoral dissertation, Lund University].

Noland, R. B., Sinclair, J. A., Klein, N. J., & Brown, C. (2017). How good is pedestrian fatality data? *Journal of Transport & Health*, *7*, 3–9.

Noor, F., Ashfaq, U. A., Bakar, A., Allemailem, K. S., Alharbi, B. F., Al-Megrin, W. A. I., Tahir ul Qamar, M., et al. (2023). Discovering common pathogenic processes between covid-19 and hfrs by integrating rna-seq differential expression analysis with machine learning. *Frontiers in Microbiology*, *14*, 1175844.

Obeng, K., & Rokonuzzaman, M. (2013). Pedestrian injury severity in automobile crashes.

O'Hern, S., Stephan, K., Qiu, J., & Oxley, J. (2019). A simulator study of driving behavior and mental workload in mixed-use arterial road environments. *Traffic injury prevention*, *20*(6), 648–654.

Oña López, J. J. D., López Maldonado, G., Abellán Mulero, J., et al. (2013). Extracting decision rules from police accident reports through decision trees.

Onelcin, P., & Alver, Y. (2017). The crossing speed and safety margin of pedestrians at signalized intersections. *Transportation Research Procedia*, *22*, 3–12.

Onieva-García, M. Á., Martínez-Ruiz, V., Lardelli-Claret, P., Jiménez-Moleón, J. J., Amezcua-Prieto, C., de Dios Luna-del-Castillo, J., & Jiménez-Mejías, E. (2016). Gender and age differences in components of traffic-related pedestrian death rates: Exposure, risk of crash and fatality rate. *Injury epidemiology*, *3*(1), 1–10.

Organization, W. H. (2003). *The world health report 2003: Shaping the future*.

Organization, W. H. (2013). *Who global status report on road safety 2013*.

Organization, W. H. (2015). *Global status report on road safety 2015*.

Organization, W. H. (2018). *Global status report on road safety 2018*.

Öström, M., & Eriksson, A. (2001). Pedestrian fatalities and alcohol. *Accident Analysis & Prevention*, *33*(2), 173–180.

Oviedo-Trespalacios, O., & Scott-Parker, B. (2017). Footbridge usage in high-traffic flow highways: The intersection of safety and security in pedestrian decision-making. *Transportation research part F: traffic psychology and behaviour*, *49*, 177–187.

Owen, N., Humpel, N., Leslie, E., Bauman, A., & Sallis, J. F. (2004). Understanding environmental influences on walking: Review and research agenda. *American journal of preventive medicine*, *27*(1), 67–76.

Oxley, J., Fildes, B., Ihsen, E., Charlton, J., & Day, R. (1997). Differences in traffic judgements between young and old adult pedestrians. *Accident Analysis & Prevention*, *29*(6), 839–847.

Oxley, J., Charlton, J. L., & Fildes, B. (2005). *The effect of cognitive impairment on older pedestrian behaviour and crash risk*. Monash University Accident Research Centre Melbourne.

Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., & Yannis, G. (2019–Accessed 24.09.2018). Review and ranking of crash risk factors related to the road infrastructure.

Park, S.-H., & Bae, M.-K. (2020). Exploring the determinants of the severity of pedestrian injuries by pedestrian age: A case study of daegu metropolitan city, south korea. *International journal of environmental research and public health*, *17*(7), 2358.

Pasanen, E. (1992). *Driving speeds and pedestrian safety: A mathematical model* (tech. rep.).

Peng, C., & Xu, C. (2022). Combined variable speed limit and lane change guidance for secondary crash prevention using distributed deep reinforcement learning. *Journal of Transportation Safety & Security*, *14*(12), 2166–2191.

PIARC. (2015). World road association activity report 2012-2015.

Pilkington, P., & Kinra, S. (2005). Effectiveness of speed cameras in preventing road traffic collisions and related casualties: Systematic review. *Bmj*, *330*(7487), 331–334.

PRACT. (2016). Prediction models and crash modification factors.

Prasetijo, J., & Ahmad, H. (2012). *Effectiveness of conflict method for capacity estimation of nonpriority intersections* (tech. rep.).

Prato, C. G., Bekhor, S., Galtzur, A., Mahalel, D., & Prashker, J. (2010). Exploring the potential of data mining techniques for the analysis of accident patterns. *12th World Conference on Transportation Research*.

Prato, C. G., Gitelman, V., & Bekhor, S. (2012). Mapping patterns of pedestrian fatal accidents in israel. *Accident Analysis & Prevention*, *44*(1), 56–62.

Premebida, C., Ludwig, O., & Nunes, U. (2009). Lidar and vision-based pedestrian detection system. *Journal of Field Robotics*, *26*(9), 696–711.

Programme, I. R. A. (2015). Vaccines for roads: Third edition.

Publishing, O., & Forum, I. T. (2008). *Towards zero: Ambitious road safety targets and the safe system approach*. Organisation for Economic Co-operation; Development.

Raifman, M., Lambert, K. F., Levy, J. I., & Kinney, P. L. (2021). Mortality implications of increased active mobility for a proposed regional transportation emission cap-and-invest program. *Journal of Urban Health*, 1–13.

Rankavat, S., & Tiwari, G. (2016). Pedestrians perceptions for utilization of pedestrian facilities–delhi, india. *Transportation research part F: traffic psychology and behaviour*, *42*, 495–499.

Ren, G., Zhou, Z., Wang, W., Zhang, Y., & Wang, W. (2011). Crossing behaviors of pedestrians at signalized intersections: Observational study and survey in china. *Transportation research record*, *2264*(1), 65–73.

Robbins, S. (2023). The many meanings of meaningful human control. *AI and Ethics*, 1–12.

Rouse, J. B., & Schwebel, D. C. (2019). Supervision of young children in parking lots: Impact on child pedestrian safety. *Journal of safety research*, *70*, 201–206.

Ryser, L., & Halseth, G. (2008). Institutional barriers to incorporating climate responsive design in commercial redevelopment. *Environment and Planning B: Planning and Design*, *35*(1), 34–55.

Ryser, L., Halseth, G., Markey, S., Gunton, C., & Argent, N. (2019). Path dependency or investing in place: Understanding the changing conditions for rural resource regions. *The Extractive Industries and Society*, *6*(1), 29–40.

Sabrina, F., Sohail, S., Thakur, S., Azad, S., & Wasimi, S. (2020). Use of deep learning approach on uav imagery to detect mistletoe infestation. *2020 IEEE Region 10 Symposium (TENSYMP)*, 556–559.

Saha, S., Roy, J., Pradhan, B., & Hembram, T. K. (2021). Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at east sikkim himalayan, india. *Advances in Space Research*, *68*(7), 2819–2840.

Sakai, T., Onishi, H., & Miyazoe, K. (2015). Evaluation of optimal threshold of z score map for statistical brain image analysis. *Nihon Hoshasen Gijutsu Gakkai Zasshi*, *71*(4), 325–331.

Salmon, P. M., Read, G. J., & Stevens, N. J. (2016). Who is in control of road safety? a stamp control structure analysis of the road transport system in queensland, australia. *Accident Analysis & Prevention*, *96*, 140–151.

Sârbescu, P., Stanojević, P., & Jovanović, D. (2014). A cross-cultural analysis of aggressive driving: Evidence from serbia and romania. *Transportation research part F: traffic psychology and behaviour*, *24*, 210–217.

Sarkar, S., Tay, R., & Hunt, J. D. (2011). Logistic regression model of risk of fatality in vehicle–pedestrian crashes on national highways in bangladesh. *Transportation research record*, *2264*(1), 128–137.

Sarraf, R., & McGuire, M. P. (2020). Integration and comparison of multi-criteria decision making methods in safe route planner. *Expert Systems with Applications*, *154*, 113399.

Sasidharan, L., & Donnell, E. T. (2014). Propensity scores-potential outcomes framework to incorporate severity probabilities in the highway safety manual crash prediction algorithm. *Accident Analysis & Prevention*, *71*, 183–193.

Savin-Baden, M., & Major, C. H. (2023). *Qualitative research: The essential guide to theory and practice*. Taylor & Francis.

Schlögl, M., & Stütz, R. (2019). Methodological considerations with data uncertainty in road safety analysis. *Accident Analysis & Prevention*, *130*, 136–150.

Schmidt, C. O., Struckmann, S., Enzenbach, C., Reineke, A., Stausberg, J., Damerow, S., Huebner, M., Schmidt, B., Sauerbrei, W., & Richter, A. (2021). Facilitating harmonized data quality assessments. a data quality framework for observational health research data collections with software implementations in r. *BMC medical research methodology*, *21*(1), 1–15.

Schuurman, N., Cinnamon, J., Crooks, V. A., & Hameed, S. M. (2009). Pedestrian injury and the built environment: An environmental scan of hotspots. *BMC public health*, *9*(1), 1–10.

Senserrick, T., Boufous, S., De Rome, L., Ivers, R., & Stevenson, M. (2014). Detailed analysis of pedestrian casualty collisions in victoria, australia. *Traffic injury prevention*, *15*(sup1), S197–S205.

Serna, A., & Marcotegui, B. (2014). Detection, segmentation and classification of 3d urban objects using mathematical morphology and supervised learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, 243–255.

Shaheed, M. S. B., Gkritza, K., Zhang, W., & Hans, Z. (2013). A mixed logit analysis of two-vehicle crash severities involving a motorcycle. *Accident Analysis & Prevention*, *61*, 119–128.

Shapiro, A. (2020). 'embodiments of the invention': Patents and urban diagrammatics in the smart city. *Convergence*, *26*(4), 751–774.

Shirkhorshidi, A. S., Aghabozorgi, S., & Wah, T. Y. (2015). A comparison study on similarity and dissimilarity measures in clustering continuous data. *PloS one*, *10*(12), e0144059.

Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *Journal of traffic and transportation engineering (English edition)*, *7*(6), 775–790.

Soh, J., Singh, P., Soh, J., & Singh, P. (2020). Machine learning operations. *Data Science Solutions on Azure: Tools and Techniques Using Databricks and MLOps*, 259–279.

Soilán, M., Riveiro, B., Sánchez-Rodríguez, A., & Arias, P. (2018). Safety assessment on pedestrian crossing environments using mls data. *Accident Analysis & Prevention*, *111*, 328–337.

Star rating for all road users. (2021). *International Road Access Programme.*

Stefanova, T. (2015). *Factors shaping pedestrians' unsafe behaviour at actively protected level crossings* [Doctoral dissertation, Paris 8].

Stigson, H. (2009). *A safe road transport system: Factors influencing injury outcome for car occupants.* Institutionen för klinisk neurovetenskap/Department of Clinical Neuroscience.

Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., Andersen, H., Fuchs, J., Kim, M., Petit, J.-E., & Favez, O. (2020). Meteorology-driven variability of air pollution (pm1) revealed with explainable machine learning. *Atmos. Chem. Phys. Discuss*, *2020*, 1–35.

Stone, J. R., Chae, K., & Pillalamarri, S. (2002). *The effects of roundabouts on pedestrian safety.* Southeastern Transportation Center.

Sullivan, J. M., & Flannagan, M. J. (2002). The role of ambient light level in fatal crashes: Inferences from daylight saving time transitions. *Accident Analysis & Prevention*, *34*(4), 487–498.

Syazwan, S. M., Baba, M. D., Zarifah, H. N., Hafeez, A. A., & Faradila, P. N. (2017). Prevalence of distracted pedestrians while crossing: A study of malaysia's situation. *MATEC Web of Conferences*, *90*, 01031.

Tabone, W., & De Winter, J. (2023). Using chatgpt for human–computer interaction research: A primer. *Manuscript submitted for publication.*

Takahashi, Y., Ueki, M., Yamada, M., Tamiya, G., Motoike, I. N., Saigusa, D., Sakurai, M., Nagami, F., Ogishima, S., Koshiba, S., et al. (2020). Improved metabolomic data-based prediction of depressive symptoms using nonlinear machine learning with feature selection. *Translational psychiatry*, *10*(1), 1–12.

Tanur, J. M. (2011). Margin of error. In *International encyclopedia of statistical science* (pp. 765–765). Springer.

Tay, R. (2003). Effectiveness of road safety advertising in reducing fatal crashes in new zealand. *Australia and New Zealand Marketing Academy Conference*, 1–3.

Tesema, T. B., Abraham, A., & Grosan, C. (2005). Rule mining and classification of road traffic accidents using adaptive regression trees. *International Journal of Simulation*, *6*(10), 80–94.

TfL. (2018). Transport for london – casualties in greater london during 2017 fact-sheet — accessed 21.03.2023.

TfL. (2019). Transport for london – casualties in greater london during 2017 fact-sheet — accessed 21.03.2023.

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, *72*, 244–256.

Torbic, D. J., Harwood, D. W., Bokenkroger, C. D., Srinivasan, R., Carter, D., Zegeer, C. V., & Lyon, C. (2010). Pedestrian safety prediction methodology for urban signalized intersections. *Transportation research record*, *2198*(1), 65–74.

Törnros, J. E., & Bolling, A. K. (2005). Mobile phone use—effects of handheld and handsfree phones on driving performance. *Accident Analysis & Prevention*, *37*(5), 902–909.

Tournier, I., Dommes, A., & Cavallo, V. (2016). Review of safety and mobility issues among older pedestrians. *Accident Analysis & Prevention*, *91*, 24–35.

Treno, A. J., Marzell, M., Gruenewald, P. J., & Holder, H. (2014). A review of alcohol and other drug control policy research. *Journal of Studies on Alcohol and Drugs, Supplement*, (s17), 98–107.

Tripathy, A. K., Mishra, A. K., & Das, T. K. (2017). Smart lighting: Intelligent and weather adaptive lighting in street lights using iot. *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 1236–1239.

Trueck, S., & Rachev, S. T. (2009). *Rating based modeling of credit risk: Theory and application of migration matrices*. Academic press.

Tsigdinos, S., Tzouras, P. G., Bakogiannis, E., Kepaptsoglou, K., & Nikitas, A. (2022). The future urban road: A systematic literature review-enhanced q-method study with experts. *Transportation Research Part D: Transport and Environment*, *102*, 103158.

Tung, Y.-C., Liu, Y.-C., & Ou, Y.-K. (2008). The pedestrian road-crossing behaviors between older and younger age groups. *Proceedings of the 9th Asia Pacific Industrial Engineering & Management Systems Conference, APIEMS*, 6–12.

Turner, S., Singh, R., & Nates, G. (2012). Crash prediction models for signalised intersections: Signal phasing and geometry june 2012.

Turner, S., Sener, I. N., Martin, M. E., Das, S., Hampshire, R. C., Fitzpatrick, K., Molnar, L. J., Colety, M., Robinson, S., Shipp, E., et al. (2017). *Synthesis of methods for estimating pedestrian and bicyclist exposure to risk at areawide levels and on specific transportation facilities* (tech. rep.). United States. Federal Highway Administration. Office of Safety.

Twisk, D., Davidse, R., & Schepers, P. (2015). Challenges in reducing bicycle casualties with high volume cycle use: Lessons from the netherlands. *Cycling Futures: From Research into Practice. Routledge*, 137–154.

Twisk, D. A., Commandeur, J. J., Vlakveld, W. P., Shope, J. T., & Kok, G. (2015). Relationships amongst psychological determinants, risk behaviour, and road crashes of young adolescent pedestrians and cyclists: Implications for road safety education programmes. *Transportation research part F: traffic psychology and behaviour*, *30*, 45–56.

Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. *Water Research*, *229*, 119422.

Urie, Y., Velaga, N. R., & Maji, A. (2016). Cross-sectional study of road accidents and related law enforcement efficiency for 10 countries: A gap coherence analysis. *Traffic injury prevention*, *17*(7), 686–691.

Useche, S. A., Hezaveh, A. M., Llamazares, F. J., & Cherry, C. (2021). Not gendered… but different from each other? a structural equation model for explaining risky road behaviors of female and male pedestrians. *Accident Analysis & Prevention*, *150*, 105942.

Uttley, J., & Fotios, S. (2017). Using the daylight savings clock change to show ambient light conditions significantly influence active travel. *Journal of Environmental Psychology*, *53*, 1–10.

Uzan, G., & Wagstaff, P. (2018). Solid: A model to analyse the accessibility of transport systems for visually impaired people. In *Mobility of visually impaired people* (pp. 353–373). Springer.

Viano, D. C., & Parenteau, C. S. (2021). Change in seating position of children in towaway crashes from 1989 to 2019. *Accident Analysis & Prevention*, *159*, 106280.

Vignali, V., Pazzini, M., Ghasemi, N., Lantieri, C., Simone, A., & Dondi, G. (2020). The safety and conspicuity of pedestrian crossing at roundabouts: The effect of median refuge island and zebra markings. *Transportation research part F: traffic psychology and behaviour*, *68*, 94–104.

Vijayawargiya, V., & Rokade, S. (2017). Identification of factors affecting pedestrian level of service of crosswalks at roundabouts. *International Research Journal of Engineering and Technology (IRJET)*, *4*(6), 342–346.

Wang, J., Kong, Y., & Fu, T. (2019). Expressway crash risk prediction using back propagation neural network: A brief investigation on safety resilience. *Accident Analysis & Prevention*, *124*, 180–192.

Wedagama, D. M. P., & Dissanayake, D. (2020). A model of latent class multinomial logit to investigate motorcycle accident injuries. *Engineering and Applied Science Research*, *47*(4), 422–429.

Wennberg, H., Ståhl, A., & Hydén, C. (2009). Older pedestrians' perceptions of the outdoor environment in a year-round perspective. *European Journal of Ageing*, *6*(4), 277–290.

Wiljer, D., & Hakim, Z. (2019). Developing an artificial intelligence–enabled health care practice: Rewiring health care professions for better care. *Journal of medical imaging and radiation sciences*, *50*(4), S8–S14.

Wilson, D. G., & Grayson, G. (1980). *Age-related differences in the road crossing behaviour of adult pedestrians* (tech. rep.).

Wisch, M., Hellmann, A., Lerner, M., Hierlinger, T., Labenski, V., Wagner, M., Feifel, H., Robescu, O., Renoux, P., & Groult, X. (2019). Car-to-car accidents at intersections in europe and identification of use cases for the test and assessment of respective active vehicle safety systems. *26th International Technical Conference on the Enhanced Safety of Vehicles, Eindhoven, Netherlands*, 10–13.

Wisdom, J., & Creswell, J. W. (2013). Mixed methods: Integrating quantitative and qualitative data collection and analysis while studying patient-centered medical home models. *Rockville: Agency for Healthcare Research and Quality.*

Witten, I. (n.d.). H., frank, e., hall, m., a., 2011. *Data Mining: Practical Machine Learning Tools and Techniques.*

Wu, P., Meng, X., Song, L., & Zuo, W. (2019). Crash risk evaluation and crash severity pattern analysis for different types of urban junctions: Fault tree analysis and association rules approaches. *Transportation research record*, *2673*(1), 403–416.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), 1–37.

Xiupeng, S. (2019). Accident risk assessment and prediction using surrogate indicators and machine learning.

Yannis, G., & Karlaftis, M. G. (2010). Weather effects on daily traffic accidents and fatalities: A time series count data approach. *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, *10*, 14.

Yannis, G., Kondyli, A., & Mitzalis, N. (2013). Effect of lighting on frequency and severity of road accidents. *Proceedings of the Institution of Civil Engineers-Transport*, *166*(5), 271–281.

Yong, Y., Chongxun, Z., & Pan, L. (2004). A novel fuzzy c-means clustering algorithm for image thresholding. *Measurement science review*, *4*(1), 11–19.

Yuan, W., Fu, R., Ma, Y., & Guo, Y.-S. (2011). Effects of vehicle speed and traffic sign text height on drivers' visual search patterns. *Journal of traffic and transportation engineering (Xi'an, Shaanxi)*, *11*(1), 119–126.

Zahid, M., Chen, Y., Jamal, A., Al-Ofi, K. A., & Al-Ahmadi, H. M. (2020). Adopting machine learning and spatial analysis techniques for driver risk assessment: Insights from a case study. *International journal of environmental research and public health*, *17*(14), 5193.

Zangenehpour, S., Miranda-Moreno, L. F., & Saunier, N. (2015). Automated classification based on video data at intersections with heavy pedestrian and bicycle traffic: Methodology and application. *Transportation research part C: emerging technologies*, *56*, 161–176.

Zegeer, C. V., & Bushell, M. (2012). Pedestrian crash trends and potential countermeasures from around the world. *Accident Analysis & Prevention*, *44*(1), 3–11.

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 121–130.

Zeng, W., Chen, P., Nakamura, H., & Iryo-Asano, M. (2014). Application of social force model to pedestrian behavior analysis at signalized crosswalk. *Transportation research part C: emerging technologies*, *40*, 143–159.

Zhang, Y., Jing, L., Sun, C., Fang, J., & Feng, Y. (2019). Human factors related to major road traffic accidents in china. *Traffic injury prevention*, *20*(8), 796–800.

Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., & Wang, Z. (2019). Traffic accident's severity prediction: A deep-learning approach-based cnn network. *IEEE Access*, *7*, 39897–39910.

Zheng, T., Qu, W., Ge, Y., Sun, X., & Zhang, K. (2017). The joint effect of personality traits and perceived stress on pedestrian behavior in a chinese sample. *PloS one*, *12*(11), e0188153.

Zhu, D., Sze, N., & Bai, L. (2021). Roles of personal and environmental factors in the red light running propensity of pedestrian: Case study at the urban crosswalks. *Transportation research part F: traffic psychology and behaviour*, *76*, 47–58.

Zhu, M., Zhao, S., Coben, J. H., & Smith, G. S. (2012). Why more male pedestrians die in vehicle-pedestrian collisions than females: A decompositional analysis.

A. Appendix

A.1. Status 19 Form

A.2. UK Highway Local Authorities by Region

A.3. FCM Deployment

A.1.1. Road Attributes

Sept 2011

MG NSRF/A

ACCIDENT STATISTICS

Incident URN

Other ref.

1.3 ACCIDENT REFERENCE

*FATAL / SERIOUS / SLIGHT

1.9 *TIME* H H M M

*DAY** Su M T W Th F S

1.7 *DATE* D D M M 2 0 Y Y

1st Road Class & No. *or (Unclassified - UC)* *(Not Known - NK)*

1st Road Name

Outside House No. *or Name or Marker Post No.*

at junction with / or *metres* N S E W ** of*

2nd Road Class & No. *or (Unclassified - UC)* *(Not Known - NK)*

2nd Road Name

Town

Sector /Beat No.

County or Borough

Parish No. or Name

1.10 *Local Auth No. (if known)*

1.11 *Grid Reference* E — N

REPORTING OFFICER *Name*

Number

BCU/Stn

1.2 *Force*

Tel Number

1.5	Number of vehicles

1.6	Number of casualties

1.14 ROAD TYPE		✗
Roundabout	1	
One way street	2	
Dual carriageway	3	
Single carriageway	6	
Slip road	7	
Unknown	9	

1.15	Speed Limit (Permanent)

1.16 JUNCTION DETAIL		✗
Not at or within 20 metres of junction	00	
Roundabout	01	
Mini roundabout	02	
T or staggered junction	03	
Slip road	05	
Crossroads	06	
Junction more than four arms (not RAB)	07	
Using private drive or entrance	08	
Other junction	09	

JUNCTION ACCIDENTS ONLY

1.17 JUNCTION CONTROL		✗
Authorised person	1	
Automatic traffic signal	2	
Stop sign	3	
Give way or uncontrolled	4	

1.20a PEDESTRIAN CROSSING - HUMAN CONTROL		✗
None within 50 metres	0	
Control by school crossing patrol	1	
Control by other authorised person	2	

1.20b PEDESTRIAN CROSSING - PHYSICAL FACILITIES		✗
No physical crossing facility within 50m	0	
Zebra crossing	1	
Pelican, puffin, toucan or similar non-junction pedestrian light crossing	4	
Pedestrian phase at traffic signal junction	5	
Footbridge or subway	7	
Central refuge — no other controls	8	

1.22 WEATHER		✗
Fine without high winds	1	
Raining without high winds	2	
Snowing without high winds	3	
Fine with high winds	4	
Raining with high winds	5	
Snowing with high winds	6	
Fog or mist — if hazard	7	
Other	8	
Unknown	9	

1.23 ROAD SURFACE CONDITION		✗
Dry	1	
Wet / Damp	2	
Snow	3	
Frost / Ice	4	
Flood (surface water over 3cm deep)	5	

1.21 LIGHT CONDITIONS		✗
Daylight:	1	
Darkness: street lights present and lit	4	
Darkness: street lights present but unlit	5	
Darkness: no street lighting	6	
Darkness: street lighting unknown	7	

1.24 SPECIAL CONDITIONS AT SITE		✗
None	0	
Auto traffic signal out	1	
Auto traffic signal partially defective	2	
Permanent road signing or marking defective or obscured	3	
Roadworks	4	
Road surface defective	5	
Oil or diesel	6	
Mud	7	

1.25 CARRIAGEWAY HAZARDS		✗
None	0	
Dislodged vehicle load in carriageway	1	
Other object in carriageway	2	
Involvement with previous accident	3	
Pedestrian in carriageway - not injured	6	
Any animal in carriageway (except ridden horse)	7	

1.26 Did a police officer attend the scene and obtain the details for this report?		✗
Yes	1	
No	2	

Subject to local directions, boxes with a grey background need not be completed if already recorded

** Circle as appropriate*

UNCLASSIFIED

MG NSRF/B

VEHICLE RECORD

Sept 2011

2.26 VEHICLE REGISTRATION MARK

Vehicle 001	
Vehicle 002	
Vehicle 003	
Vehicle 004	

2.35 WAS THE VEHICLE LEFT HAND DRIVE ✗

		VEHICLE 1	2	3	4
No	1				
Yes	2				

2.5 / 2.5a TYPE OF VEHICLE ✗

		1	2	3	4
Car	09				
Taxi / Private hire car	08				
Van - Goods vehicle 3.5 tonnes mgw and under	19				
Goods vehicle over 3.5 tonnes mgw and under 7.5 tonnes mgw	20				
Goods vehicle 7.5 tonnes mgw & over	21				
Goods vehicle - unknown weight	98				
M/cycle 50cc and under	02				
M/cycle over 50cc and up to 125cc	03				
M/cycle over 125cc and up to 500cc	04				
Motorcycle over 500cc	05				
Motorcycle - cc unknown	97				
Electric Motorcycle	23				
Pedal cycle	01				
Bus or coach (17 or more passenger seats)	11				
Minibus (8-16 passenger seats)	10				
Agricultural vehicle (include diggers etc)	17				
Ridden horse	16				
Mobility scooter	22				
Tram / Light rail	18				
Other vehicle 1	90				
2	90				
3	90				
4	90				

2.6 TOWING AND ARTICULATION ✗

		1	2	3	4
No tow or articulation	0				
Articulated vehicle	1				
Double or multiple trailer	2				
Caravan	3				
Single trailer	4				
Other tow	5				

2.22 AGE OF DRIVER (Estimate if necessary)

Vehicle 001		Vehicle 002	
Vehicle 003		Vehicle 004	

2.27 DRIVER HOME POSTCODE
or Code: 1- Unknown 2- Non UK Resident 3 - Parked & unattended

Vehicle 001		
Vehicle 002		
Vehicle 003		
Vehicle 004		

2.23 BREATH TEST ✗

		VEHICLE 1	2	3	4
Not applicable	0				
Positive	1				
Negative	2				
Not requested	3				
Refused to provide	4				
Driver not contacted at time of col'	5				
Not provided (medical reasons)	6				

2.24 HIT AND RUN ✗

		1	2	3	4
Not hit and run	0				
Hit and run	1				
Non-stop vehicle, not hit	2				

2.21 SEX OF DRIVER ✗

		1	2	3	4
Male	1				
Female	2				
Not known	3				

2.9 VEHICLE LOCATION AT TIME OF ACCIDENT RESTRICTED LANE/AWAY FROM MAIN C'WAY ✗

		1	2	3	4
On main carriageway not in restricted lane	00				
Tram / Light rail track	01				
Bus lane	02				
Busway (inc. guided busway)	03				
Cycle lane (on main carriageway)	04				
Cycleway or shared use footway (not part of main carriageway)	05				
On lay-by / hard shoulder	06				
Entering lay-by/ hard shoulder	07				
Leaving lay-by / hard shoulder	08				
Footway (pavement)	09				

2.10 JUNCTION LOCATION OF VEHICLE ✗

		1	2	3	4
Not at or within 20m of junction	0				
Approaching junction or waiting /parked at junction approach	1				
Cleared junction or waiting/ parked at junction exit	2				
Leaving roundabout	3				
Entering roundabout	4				
Leaving main road	5				
Entering main road	6				
Entering from slip road	7				
Mid junction– on roundabout or on main road	8				

2.7 MANOEUVRES ✗

		1	2	3	4
Reversing	01				
Parked	02				
Waiting to go ahead but held up	03				
Slowing or stopping	04				
Moving off	05				
U turn	06				
Turning left	07				
Waiting to turn left	08				
Turning right	09				
Waiting to turn right	10				
Changing lane to left	11				
Changing lane to right	12				
O'taking moving veh on its offside	13				
O'taking stationary veh on its offside	14				
Overtaking on nearside	15				
Going ahead left hand bend	16				
Going ahead right hand bend	17				
Going ahead other	18				

2.11 SKIDDING AND OVERTURNING ✗

		VEHICLE 1	2	3	4
No skidding, jack-knifing or overturning	0				
Skidded	1				
Skidded and overturned	2				
Jack - knifed	3				
Jack - knifed and overturned	4				
Overturned	5				

2.12 HIT OBJECT IN CARRIAGEWAY ✗

		1	2	3	4
None	00				
Previous accident	01				
Roadworks	02				
Parked vehicle	04				
Bridge - roof	05				
Bridge - side	06				
Bollard / Refuge	07				
Open door of vehicle	08				
Central island of roundabout	09				
Kerb	10				
Any animal (except ridden horse)	12				
Other object	11				

2.13 VEHICLE LEAVING CARRIAGEWAY ✗

		1	2	3	4
Did not leave carriageway	0				
Left carriageway nearside	1				
Left carriageway nearside and rebounded	2				
Left carriageway straight ahead at junction	3				
Left carriageway offside onto central reservation	4				
Left carriageway offside onto central reserve and rebounded	5				
Left carriageway offside and crossed central reservation	6				
Left carriageway offside	7				
Left carriageway offside and rebounded	8				

2.14 FIRST OBJECT HIT OFF CARRIAGEWAY ✗

		1	2	3	4
None	00				
Road sign / Traffic signal	01				
Lamp post	02				
Telegraph pole / Electricity pole	03				
Tree	04				
Bus stop / Bus shelter	05				
Central crash barrier	06				
Nearside or offside crash barrier	07				
Submerged in water (completely)	08				
Entered ditch	09				
Wall or fence	11				
Other permanent object	10				

2.16 FIRST POINT OF IMPACT ✗

		1	2	3	4
Did not impact	0				
Front	1				
Back	2				
Offside	3				
Nearside	4				

2.29 JOURNEY PURPOSE OF DRIVER/RIDER ✗

		1	2	3	4
Journey as part of work	1				
Commuting to / from work	2				
Taking school pupil to/from school	3				
Pupil riding to / from school	4				
Other	5				
Not known	6				

Subject to local directions, boxes with a grey background need not be completed if already recorded

UNCLASSIFIED

MG NSRF/C Sept 2011

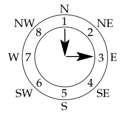
2.8 DIRECTION OF VEHICLE TRAVEL

1. Using the Example shown complete the FROM and TO boxes for the vehicles concerned, indicating direction of travel FROM and TO
2. If PARKED enter '00'

Vehicle 001		*Vehicle 002*	
FROM	TO	FROM	TO
Vehicle 003		***Vehicle 004***	
FROM	TO	FROM	TO

EXAMPLE

FROM	TO
1	3



CASUALTY RECORD

3.4 VEHICLE REFERENCE NUMBER
Enter VEH No. which CASUALTY occupied (for pedestrians, code vehicle that struck them first) e.g. 001,002 etc.

Casualty 001	0			Casualty 002	0		
Casualty 003	0			Casualty 004	0		
Casualty 005	0			Casualty 006	0		

3.18 CASUALTY HOME POSTCODE
or Code: 1- Unknown
2- Non UK Resident

Casualty 001								
Casualty 002								
Casualty 003								
Casualty 004								
Casualty 005								
Casualty 006								

3.7 SEX OF CASUALTY ✗		CASUALTY 1	2	3	4	5	6
Male	1						
Female	2						

3.8 AGE OF CASUALTY (Estimate if necessary)
For children less than a year enter 00

Casualty 001				Casualty 002			
Casualty 003				Casualty 004			
Casualty 005				Casualty 006			

3.6 CASUALTY CLASS ✗							
Driver/Rider	1						
Veh./pillion Passenger	2						
Pedestrian	3						

3.9 SEVERITY OF CASUALTY ✗							
Fatal	1						
Serious	2						
Slight	3						

3.20 CYCLE HELMET WORN ✗		CASUALTY 1	2	3	4	5	6
Not a cyclist	0						
Yes	1						
No	2						
Not known	3						

3.15 CAR PASSENGER (not driver) ✗							
Not a car passenger	0						
Front seat passenger	1						
Rear seat passenger	2						

3.16 BUS OR COACH PASSENGER ✗ (17 passenger seats or more)							
Not a bus or coach passenger	0						
Boarding	1						
Alighting	2						
Standing passenger	3						
Seated passenger	4						

LOCAL STATISTICS

3.14 SEAT BELT IN USE ✗							
Not applicable	0						
Worn and independently confirmed	1						
Worn but not independently confirmed	2						
Not worn	3						
Unknown	4						

PEDESTRIAN CASUALTIES ONLY

3.10 PEDESTRIAN LOCATION ✗		CASUALTY 1	2	3	4	5	6
In carriageway, crossing on pedestrian crossing facility	01						
In carriageway, crossing within zig-zag lines at crossing approach	02						
In carriageway, crossing within zig-zag lines at crossing exit	03						
In carriageway, crossing elsewhere within 50m of pedestrian crossing	04						
In carriageway, crossing elsewhere	05						
On footway or verge	06						
On refuge, central island or central reservation	07						
In centre of carriageway, not on refuge, island or central reservation	08						
In carriageway, not crossing	09						
Unknown or other	10						

3.12 PEDESTRIAN DIRECTION ✗		CASUALTY 1	2	3	4	5	6
Standing still	0						
Northbound	1						
Northeast bound	2						
Eastbound	3						
Southeast bound	4						
Southbound	5						
Southwest bound	6						
Westbound	7						
Northwest bound	8						
Unknown	9						

3.19 PEDESTRIAN ROAD MAINTENANCE WORKER ✗							
No / not applicable	0						
Yes	1						
Not known	2						

3.11 PEDESTRIAN MOVEMENT ✗		CASUALTY 1	2	3	4	5	6
Crossing from driver's nearside	1						
Crossing from driver's nearside-masked by parked or stationary veh'	2						
Crossing from driver's offside	3						
Crossing from driver's offside-masked by parked or stationary veh'	4						
In carriageway, stationary - not crossing (standing or playing)	5						
In carriageway, stationary -not crossing (standing or playing), masked by parked or stationary veh'	6						
Walking along in carriageway-facing traffic	7						
Walking along in carriageway-back to traffic	8						
Unknown or other	9						

UNCLASSIFIED

MG NSRF/D

RESTRICTED

CONTRIBUTORY FACTORS

Sept 2011

1. Select up to six factors from the grid, relevant to the accident.
2. Factors may be shown in any order, but an indication must be given of whether each factor is ***very likely (A)*** or ***possible (B).***
3. Only include factors that you consider contributed to the accident. (i.e. do NOT include "Poor road surface" unless relevant).
4. More than one factor may, if appropriate, be related to the same road user.
5. The same factor may be related to more than one road user.
6. The participant should be identified by the relevant vehicle or casualty ref no. (e.g. 001, 002 etc.), preceded by "V" if the factor applies to a vehicle, driver/rider or the road environment (e.g. V002), or "C" if the factor relates to a pedestrian or passenger casualty (e.g. C001).
7. Enter U000 if the factor relates to an uninjured pedestrian.

	Road Environment Contributed	**103**	**102**	**101**	**110**	**108**	**107**	**109**	**104**	**105**	**106**
		Slippery road (due to weather)	Deposit on road (e.g. oil, mud, chippings)	Poor or defective road surface	Sunken, raised or slippery inspection cover	Road layout (e.g. bend, hill, narrow carriageway)	Temporary road layout (e.g. contraflow)	Animal or object in carriageway	Inadequate or masked signs or road markings	Defective traffic signals	Traffic calming (e.g. speed cushions, road humps, chicanes)
	Vehicle Defects	**201**	**202**	**203**	**204**	**205**	**206**				
		Tyres illegal, defective or under-inflated	Defective lights or indicators	Defective brakes	Defective steering or suspension	Defective or missing mirrors	Overloaded or poorly loaded vehicle or trailer				
Driver/Rider Only (Includes Pedal Cycles and Horse Riders)	Injudicious Action	**308**	**306**	**302**	**301**	**307**	**310**	**305**	**304**	**309**	**303**
		Following too close	Exceeding speed limit	Disobeyed Give Way or Stop sign or markings	Disobeyed automatic traffic signal	Travelling too fast for conditions	Cyclist entering road from pavement	Illegal turn or direction of travel	Disobeyed pedestrian crossing facility	Vehicle travelling along pavement	Disobeyed double white lines
	Driver/ Rider Error or Reaction	**405**	**406**	**403**	**408**	**409**	**401**	**402**	**404**	**407**	**410**
		Failed to look properly	Failed to judge other person's path or speed	Poor turn or manoeuvre	Sudden braking	Swerved	Junction overshoot	Junction restart (moving off at junction)	Failed to signal or misleading signal	Too close to cyclist, horse or pedestrian	Loss of control
	Impairment or Distraction	**501**	**502**	**508**	**503**	**509**	**510**	**505**	**504**	**507**	**506**
		Impaired by alcohol	Impaired by drugs (illicit or medicinal)	Driver using mobile phone	Fatigue	Distraction in vehicle	Distraction outside vehicle	Illness or disability, mental or physical	Uncorrected, defective eyesight	Rider wearing dark clothing	Not displaying lights at night or in poor visibility
	Behaviour or Inexperience	**602**	**605**	**601**	**603**	**607**	**606**	**604**			
		Careless, reckless or in a hurry	Learner or inexperienced driver/rider	Aggressive driving	Nervous, uncertain or panic	Unfamiliar with model of vehicle	Inexperience of driving on the left	Driving too slow for conditions or slow vehicle (e.g. tractor)			
	Vision Affected by	**701**	**703**	**706**	**707**	**708**	**705**	**710**	**702**	**704**	**709**
		Stationary or parked vehicle(s)	Road layout (e.g. bend, winding road, hill crest)	Dazzling sun	Rain, sleet, snow or fog	Spray from other vehicles	Dazzling headlights	Vehicle blind spot	Vegetation	Buildings, road signs, street furniture	Visor or windscreen dirty, scratched or frosted etc.
	Pedestrian Only (Casualty or Uninjured)	**802**	**808**	**803**	**801**	**806**	**807**	**805**	**804**	**809**	**810**
		Failed to look properly	Careless, reckless or in a hurry	Failed to judge vehicle's path or speed	Crossing road masked by stationary or parked vehicle	Impaired by alcohol	Impaired by drugs (illicit or medicinal)	Dangerous action in carriageway (e.g. playing)	Wrong use of pedestrian crossing facility	Pedestrian wearing dark clothing at night	Disability or illness, mental or physical
	Special Codes	**901**	**902**	**903**	**904**						***999**
		Stolen vehicle	Vehicle in course of crime	Emergency vehicle on a call	Vehicle door opened or closed negligently						Other – Please specify below

	1st	***2nd***	***3rd***	***4th***	***5th***	***6th***
Factor in the accident						
Which participant? (e.g. V001, C001, U000)						
Very likely (A) or Possible (B)						

*If 999 Other, give brief details

(Note: Only use if another factor contributed to the accident ***and include it in the text description of how the accident occurred***)

These factors reflect the reporting officer's opinion at the time of reporting and may not be the result of extensive investigation

RESTRICTED

ENGLAND

London

E09000001	City of London
E09000002	Barking and Dagenham
E09000003	Barnet
E09000004	Bexley
E09000005	Brent
E09000006	Bromley
E09000007	Camden
E09000008	Croydon
E09000009	Ealing
E09000010	Enfield
E09000011	Greenwich
E09000012	Hackney
E09000013	Hammersmith and Fulham
E09000014	Haringey
E09000015	Harrow
E09000016	Havering
E09000017	Hillingdon
E09000018	Hounslow
E09000019	Islington
E09000020	Kensington and Chelsea
E09000021	Kingston upon Thames
E09000022	Lambeth
E09000023	Lewisham
E09000024	Merton
E09000025	Newham
E09000026	Redbridge
E09000027	Richmond upon Thames
E09000028	Southwark
E09000029	Sutton
E09000030	Tower Hamlets
E09000031	Waltham Forest
E09000032	Wandsworth
E09000033	Westminster
EHEATHROW	London Airport (Heathrow)

Cumbria

E07000026	Allerdale
E07000027	Barrow-in-Furness
E07000028	Carlisle
E07000029	Copeland
E07000030	Eden
E07000031	South Lakeland

Lancashire

E07000117	Burnley
E07000118	Chorley
E07000119	Fylde
E07000120	Hyndburn
E07000121	Lancaster
E07000122	Pendle
E07000123	Preston
E07000124	Ribble Valley
E07000125	Rossendale
E07000126	South Ribble
E07000127	West Lancashire
E07000128	Wyre

Unitary authorities

E06000008	Blackburn with Darwen
E06000009	Blackpool

Merseyside

E08000011	Knowsley
E08000012	Liverpool
E08000013	St. Helens
E08000014	Sefton
E08000015	Wirral

Greater Manchester

E08000001	Bolton
E08000002	Bury
E08000003	Manchester
E08000004	Oldham
E08000005	Rochdale
E08000006	Salford
E08000007	Stockport
E08000008	Tameside
E08000009	Trafford
E08000010	Wigan

Cheshire

E06000049	Cheshire East
E06000050	Cheshire West and Chester
E06000006	Halton
E06000007	Warrington

Northumbria

E06000048	Northumberland

Tyne and Wear

E08000020	Gateshead
E08000021	Newcastle upon Tyne
E08000022	North Tyneside
E08000023	South Tyneside
E08000024	Sunderland

Durham

E06000047	County Durham
E06000005	Darlington

North Yorkshire

E07000163	Craven
E07000164	Hambleton
E07000165	Harrogate
E07000166	Richmondshire
E07000167	Ryedale
E07000168	Scarborough
E07000169	Selby

Unitary authority

E06000014	York

West Yorkshire

E08000032	Bradford
E08000033	Calderdale
E08000034	Kirklees
E08000035	Leeds
E08000036	Wakefield

South Yorkshire

E08000016	Barnsley
E08000017	Doncaster
E08000018	Rotherham
E08000019	Sheffield

Humberside

E06000010	Kingston upon Hull, City of
E06000011	East Riding of Yorkshire
E06000013	North Lincolnshire
E06000012	North East Lincolnshire

Cleveland

E06000001	Hartlepool
E06000003	Redcar and Cleveland
E06000002	Middlesbrough
E06000004	Stockton-on-Tees

West Midlands

E08000025	Birmingham
E08000026	Coventry
E08000027	Dudley
E08000028	Sandwell
E08000029	Solihull
E08000030	Walsall
E08000031	Wolverhampton

Staffordshire

E07000192	Cannock Chase
E07000193	East Staffordshire
E07000194	Lichfield
E07000195	Newcastle-under-Lyme
E07000196	South Staffordshire
E07000197	Stafford
E07000198	Staffordshire Moorlands
E07000199	Tamworth

Unitary authority

E06000021	Stoke-on-Trent

West Mercia

Worcestershire

E07000234	Bromsgrove
E07000235	Malvern Hills
E07000236	Redditch
E07000237	Worcester
E07000238	Wychavon
E07000239	Wyre Forest

Unitary authorities

E06000019	Herefordshire, County of
E06000051	Shropshire
E06000020	Telford and Wrekin

Warwickshire

E07000218	North Warwickshire
E07000219	Nuneaton and Bedworth
E07000220	Rugby
E07000221	Stratford-upon-Avon
E07000222	Warwick

Derbyshire

E07000032	Amber Valley
E07000033	Bolsover
E07000034	Chesterfield
E07000036	Erewash
E07000037	High Peak
E07000038	North East Derbyshire
E07000039	South Derbyshire
E07000035	Derbyshire Dales

Unitary authority

E06000015	Derby

Nottinghamshire

E07000170	Ashfield
E07000171	Bassetlaw
E07000172	Broxtowe
E07000173	Gedling
E07000174	Mansfield
E07000175	Newark and Sherwood
E07000176	Rushcliffe

Unitary authority

E06000018	Nottingham

Lincolnshire

E07000136	Boston
E07000137	East Lindsey
E07000138	Lincoln
E07000139	North Kesteven
E07000140	South Holland
E07000141	South Kesteven
E07000142	West Lindsey

Leicestershire

E07000129	Blaby
E07000132	Hinckley and Bosworth
E07000130	Charnwood
E07000131	Harborough
E07000133	Melton
E07000134	North West Leicestershire
E07000135	Oadby and Wigston

Unitary authorities

E06000016	Leicester
E06000017	Rutland

Northamptonshire

E07000150	Corby
E07000151	Daventry
E07000152	East Northamptonshire
E07000153	Kettering
E07000154	Northampton
E07000155	South Northamptonshire
E07000156	Wellingborough

Cambridgeshire

E07000008	Cambridge
E07000009	East Cambridgeshire
E07000010	Fenland
E07000011	Huntingdonshire
E07000012	South Cambridgeshire

Unitary authority

E06000031	Peterborough

Norfolk

E07000143	Breckland
E07000144	Broadland
E07000145	Great Yarmouth
E07000148	Norwich
E07000147	North Norfolk
E07000149	South Norfolk
E07000146	King's Lynn and West Norfolk

Suffolk

E07000200	Babergh
E07000201	Forest Heath
E07000202	Ipswich
E07000203	Mid Suffolk
E07000204	St. Edmundsbury
E07000205	Suffolk Coastal
E07000206	Waveney

Bedfordshire

E06000055	Bedford
E06000032	Luton
E06000056	Central Bedfordshire

Hertfordshire

E07000095	Broxbourne
E07000096	Dacorum
E07000097	East Hertfordshire
E07000098	Hertsmere
E07000099	North Hertfordshire
E07000100	St. Albans
E07000101	Stevenage
E07000102	Three Rivers
E07000103	Watford
E07000104	Welwyn Hatfield

Essex

E07000066	Basildon
E07000067	Braintree
E07000068	Brentwood
E07000069	Castle Point
E07000070	Chelmsford
E07000071	Colchester
E07000072	Epping Forest
E07000073	Harlow
E07000074	Maldon
E07000075	Rochford
E07000076	Tendring
E07000077	Uttlesford

Unitary authorities

E06000033	Southend-on-Sea
E06000034	Thurrock

Thames Valley Police

Unitary authorities (ex Berkshire)

E06000036	Bracknell Forest
E06000037	West Berkshire
E06000038	Reading
E06000039	Slough
E06000040	Windsor and Maidenhead
E06000041	Wokingham

Buckinghamshire

E07000004	Aylesbury Vale
E07000006	South Bucks
E07000005	Chiltern
E07000007	Wycombe

Unitary authority

E06000042	Milton Keynes

Oxfordshire

E07000177	Cherwell
E07000178	Oxford
E07000180	Vale of White Horse
E07000179	South Oxfordshire
E07000181	West Oxfordshire

Hampshire

E07000084	Basingstoke and Deane
E07000086	Eastleigh
E07000087	Fareham
E07000088	Gosport
E07000089	Hart
E07000090	Havant
E07000091	New Forest
E07000085	East Hampshire
E07000092	Rushmoor
E07000093	Test Valley
E07000094	Winchester

Unitary authorities

E06000046	Isle of Wight
E06000044	Portsmouth
E06000045	Southampton

Suffolk

E07000200	Babergh
E07000201	Forest Heath
E07000202	Ipswich
E07000203	Mid Suffolk
E07000204	St. Edmundsbury
E07000205	Suffolk Coastal
E07000206	Waveney

Bedfordshire

E06000055	Bedford
E06000032	Luton
E06000056	Central Bedfordshire

Hertfordshire

E07000095	Broxbourne
E07000096	Dacorum
E07000097	East Hertfordshire
E07000098	Hertsmere
E07000099	North Hertfordshire
E07000100	St. Albans
E07000101	Stevenage
E07000102	Three Rivers
E07000103	Watford
E07000104	Welwyn Hatfield

Essex

E07000066	Basildon
E07000067	Braintree
E07000068	Brentwood
E07000069	Castle Point
E07000070	Chelmsford
E07000071	Colchester
E07000072	Epping Forest
E07000073	Harlow
E07000074	Maldon
E07000075	Rochford
E07000076	Tendring
E07000077	Uttlesford

Unitary authorities

E06000033	Southend-on-Sea
E06000034	Thurrock

Thames Valley Police

Unitary authorities (ex Berkshire)

E06000036	Bracknell Forest
E06000037	West Berkshire
E06000038	Reading
E06000039	Slough
E06000040	Windsor and Maidenhead
E06000041	Wokingham

Buckinghamshire

E07000004	Aylesbury Vale
E07000006	South Bucks
E07000005	Chiltern
E07000007	Wycombe

Unitary authority

E06000042	Milton Keynes

Oxfordshire

E07000177	Cherwell
E07000178	Oxford
E07000180	Vale of White Horse
E07000179	South Oxfordshire
E07000181	West Oxfordshire

Hampshire

E07000084	Basingstoke and Deane
E07000086	Eastleigh
E07000087	Fareham
E07000088	Gosport
E07000089	Hart
E07000090	Havant
E07000091	New Forest
E07000085	East Hampshire
E07000092	Rushmoor
E07000093	Test Valley
E07000094	Winchester

Unitary authorities

E06000046	Isle of Wight
E06000044	Portsmouth
E06000045	Southampton

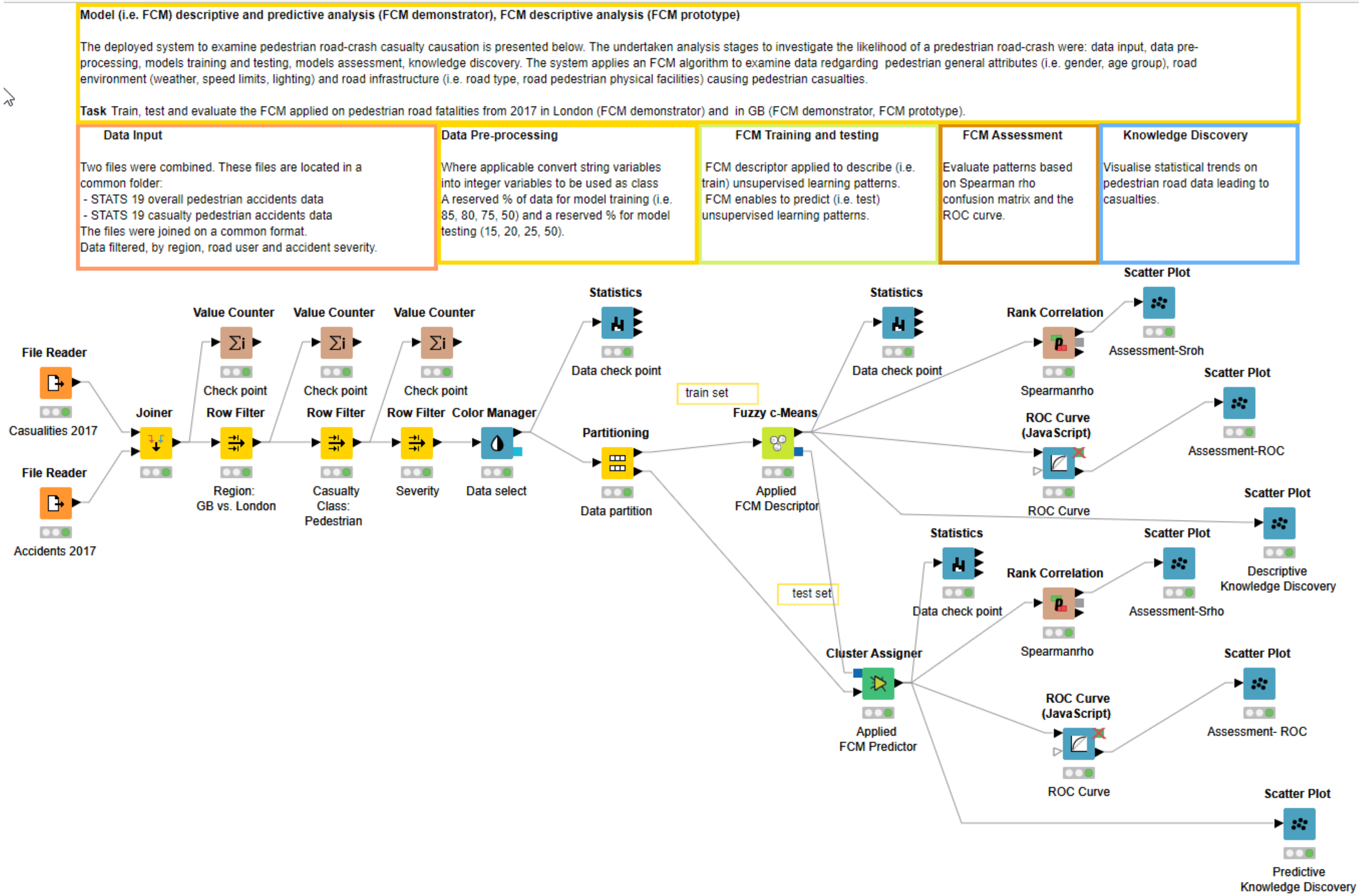
Dorset

E07000048	Christchurch
E07000050	North Dorset
E07000051	Purbeck
E07000052	West Dorset
E07000053	Weymouth and Portland
E07000049	East Dorset

Unitary authorities

E06000028	Bournemouth
E06000029	Poole

A.3.1. FCM Deployment: Spatial Analysis



A.3.2. FCM Deployment: Standalone Application