UNIVERSITY OF NAPLES FEDERICO II

Department of Civil, Architectural and Environmental Engineering

Ph.D. in Civil Systems Engineering

XXXIV Cycle

Doctoral dissertation

Ph.D. Candidate

Eng. Maria Rella Riccardi

Econometric methods and machine learning algorithms to investigate factors contributing to pedestrian crash severity

Ph.D. coordinator:

Prof. Eng. Andrea Papola

Supervisors:

Prof. Alfonso Montella

Ass. Prof. Filomena Mauriello

ABSTRACT

Road traffic crashes constitute a real concern and a serious public health problem. What is more, the worldwide burden of road traffic injuries and deaths is disproportionately borne by vulnerable road users (VRU) which include children, elderly people, pedestrians, cyclists, and motorcyclists. Reducing the increasing number of crashes involving VRU and their fatality represents the most serious challenge for the new decade of action for road safety. Among the road vulnerable users, making the second-largest group of road casualties after car occupants, pedestrians are the most susceptible to road potential risks. The severity of vehicle-pedestrian crashes further confirmed that actions to improve pedestrian safety are strongly needed to identify factors that affect (and how) crash injury severity.

The econometric models have been widely used to carry out crash severity analyses. Recently, machine learning algorithms have been used for crash severity prediction in lieu of the more traditional econometric models. To provide support for the choice of the appropriate prediction method, this research is also aimed at comparing econometric models and machine learning methods by their capability in identifying significant explanatory variables affecting crash severity and by their performances.

Analyses were carried out on three case studies using three national databases referring to the vehicle-pedestrian crashes that occurred in Great Britain in the period 2016-2018, in Sweden in the period 2015-2019, and in Italy in the period 2014-2018 to investigate how the model performances vary in presence of different sample sizes. The econometric models used in the research were the multinomial logit, the ordered logit, the random parameters multinomial logit, and the random parameters ordered logit while the machine learning algorithms include the association rules, the classification trees, the random forests, the artificial neural networks, and the support vector machine. This research further investigated the problem of imbalanced distributions of the response classes. Crash severity variable has higher variability among the severity levels' distributions which affects classification accuracy in predicting the most severe crashes of both parametric and non-parametric methods.

The quantitative models' comparison relied on the three performance metrics F-measure, G-mean, and Area Under Curve. The quantitative evaluation of the results demonstrated that machine learning tools outperformed the econometric models, and some algorithms (SVM, ANN, and RF) also prevailed over others algorithms falling under the same umbrella of machine learning tools.

The qualitative evaluation demonstrated that the machine learning tools uncover more hidden correlations among data than the econometric models and provided valuable insights on the interdependence among the several roadway, environmental, vehicle, and road users related factors contributing to the severity of pedestrian crashes. In the British case study and for fatal crashes, 19 variables were significant both in the econometric models as well as in the machine learning algorithms, 1 variable was significant only in the econometric models and 7 variables were significant only in the machine learning algorithms. In the Swedish case study, 13 variables were significant both in the econometric models as well as in the machine learning algorithms and 5 variables were only significant in the machine learning algorithms. In the Italian case study, 16 variables were significant both in the econometric models as well as in the machine learning algorithms and 3 variables were only significant in the machine learning algorithms. No further variables were identified only by the econometric models both in the Swedish and the Italian case studies. On the other hand, the random parameter econometric models provided evidence of the existence of heterogeneity among data. The presence of such variability in the effect of variables across the sample population highlights the need to account for potential unobserved heterogeneity across vehicle-pedestrian crashes as it may improve understanding and reduce erroneous inferences and predictions, producing more accurate and informative results.

In conclusion, the econometric models confirmed their advantage in offering easy to interpret outputs and understandable relations between dependent and independent variables. The magnitude of each indicator variable and its direction were clear as well. Machine learning tools, instead, exhibited higher classification accuracy and the ability to highlight more hidden relations among data. However, some machine learning tools (SVM and ANN) exhibited very high classification performances but their results are really difficult to interpret whereas, other machine learning algorithms, such as AR, CT, and RF, provided very intuitive results even though with lower prediction accuracy.

From the methodological perspective, the research results suggest that the joint use of econometric methods and machine learning algorithms may overcome the limits of each group of methods with a satisfactory trade-off between prediction accuracy and interpretation of results providing powerful insights on factors contributing to fatal and serious crashes.

From the engineering perspective, detected the interdependences between contributory patterns and severity in pedestrian crash involvement, a combination of engineering, social, and management strategies, as well as appropriate safety countermeasures, can be identified and planned to effectively moderate pedestrian crash severity, increasing the perceived safety of walking and contributing to the vision zero-deaths on road by 2050.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Mobility is an important component of the socio-economic development of individuals and society. Year by year, day by day, the transport system has become pivotal to the global economy facilitating trading, providing access to jobs and services, contributing to the development of the economy and reducing the poverty. However, despite the countless pros, the transport system is the source of many drawbacks, most of them connected with road transport. Environmental pollution, traffic jams and, most significantly, traffic crashes are the most relevant negative aspects and their direct or indirect impacts on the economy through losses of time, money and, mainly, human costs are unacceptable. Deaths and injuries resulting from road crashes represent a significant portion of the worldwide burden of ill-health and death. The number of road traffic deaths is considerably high, it has reached 1.35 million in 2016 while a significant portion of people involved in road crashes, above 50 million people, suffered non-fatal injuries with many incurring permanent disabilities because of their injury. Moreover, although low and middle-income countries have only 54% of the world's vehicles, 90% of all road traffic deaths occur in these countries. In more recent years, the rates of road crash deaths relative to the size of the world's population have stabilized. Despite this achievement, road crashes have been identified as the 8th leading cause of death for people of all ages and, what is more, they are predicted to become the 5th leading cause by 2030 (WHO, 2018), meaning that the progress observed in legislation, vehicle standards, post-care responses, has not occurred in road safety at a pace fast enough to compensate for the rising population and rapid motorization of transport taking place in many parts of the world. Furthermore, the number of pedestrian fatalities on the world's roads remains unacceptably high, with an estimated 900 pedestrians dying each day and a proportion of pedestrian fatalities of 23% (WHO, 2018). In the European Union, pedestrians account for 21% of the total road fatalities (European Commission, 2018). Pedestrian crashes are a major concern both for their number as well as for their severity. As a matter of fact, a recent study carried out in European Commission highlighted the risk of fatality for pedestrian users is nine times greater than that for occupants of four-wheeler vehicles (European Commission, 2019b).

To promote action in road safety, in September 2020, the UN General Assembly adopted the resolution A/RES/74/299 "Improving global road safety", proclaiming the Decade of Action for Road Safety 2021-2030 with the ambitious target of preventing by 2030 at least 50% of road traffic deaths and injuries caused by unsustainable transport (United Nations, 2021). Among the 17 Sustainable Development Goals (SDGs), indeed, halving the number of global deaths and injuries from road traffic crashes has been set (Goal 3.6) other than providing access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, with special attention to the needs of vulnerable road users by 2030 (Goal 11.2).

Thus, crash prevention is crucial to effectively defeat road casualties at societal, political and economic levels. Factors affecting the consequences of crashes and their costs, economic as well as social – should be adequately analysed and identified as they should be strongly considered in assessing transportation system planning and road safety management. Moreover, to improve the safety of pedestrians, the identification of the significant factors contributing to the most serious pedestrian crashes is crucial for planning, designing, and managing a safer transport system. As these factors differ from the factors affecting the severity of other crash types, properly studies focusing on vehicle-pedestrian crashes are strongly needed.

CHAPTER I ~ RESEARCH OBJECTIVES and NOVELTIES

The first chapter of this dissertation thesis provides a statement of the problem that led to the definition of the issues investigated in this research. The aims of the research are then formulated in the main objectives sections and the specific ones followed by the novelty aspects of this research. Finally, the thesis organization is provided.

1.1 Statement of the problem

Road traffic crashes constitute a real concern and a serious public health problem worldwide. Road traffic crashes are currently estimated to be the 9th leading cause of death across all age groups globally, leading to the loss of over 1.2 million lives around the world each year whereas between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability because of their injury (WHO, 2018). Thus, halving deaths and injuries resulting from a road crash represent one of the most difficult challenges to face and, with the 2030 Agenda Sustainable Development Goal targets 3.6 and 11.2, Member States aim to galvanize governments and the international community into action on road safety policy and, at the same time, to raise the share of active and environmentally sustainable transport, such as utility walking and cycling. Great emphasis on fatal and serious injuries crashes is also given in the EU Road Safety Policy Framework 2021-2030 (European Commission, 2019a, 2020), which sets a target for reducing road deaths and serious injuries in the EU by 2030 from a 2020 baseline.

Furthermore, more than half (54%) of the people who die because of a road crash are pedestrians, cyclists and motorcyclists. Pedestrians, cyclists, and motorcyclists are considered the most vulnerable road users, especially compared to vehicle occupants who have the vehicle’s protection once in a crash. Making the second largest group of road casualties after car occupants, pedestrians are the most susceptible vulnerable users in vehicle-pedestrian crashes. Indeed, the risk of fatality for pedestrian users is 7 times greater than that for occupants of 4-wheeler vehicles (European Commission, 2019b). Very recently, the European Commission published a breakdown matrix of fatalities in the EU. The matrix shows that fatalities overwhelmingly occur in crashes involving pedestrians and cyclists as road users and cars and trucks as vehicle involved (European Commission, 2021). This is, in part, a result of rapid increases in motorization followed by a road system mostly designed considering vehicular traffic’s needs neglecting vulnerable users too many times in land use planning and road traffic system design with serious consequences for their safety.

The number of pedestrian crashes recorded each year worldwide and the severity of pedestrians' outcomes are evidence that pedestrian safety is a priority task at social, political, economic and engineering levels. Safety is, indeed, an essential precondition for people to travel by using sustainable transport modes and their behaviour could change only in presence of higher perceived safety levels. Furthermore, crashes with higher injury severity and crashes involving vulnerable road users are the main cause of the greatest social costs. Thus, alleviating them will have a greater impact on the sustainability of transportation systems.

Many countries are trying to reverse these trends by promoting programs, projects and policies to support pedestrian mobility and ensure safety for all road users. In this respect, United Nations resolution 70/260, meant to improve global road safety, identifies pedestrian safety as one of the essential aspects of road safety (United Nations, 2016). The Council of the European Union (8 June 2017) states that the Member States will undertake to include cyclists and pedestrians in mobility plans, developing policies aimed at improving the safety of vulnerable users and considering, where is possible, the inclusion of a dedicated infrastructure (European Commission, 2017). To improve pedestrian safety and reach the target, the EU designed the strategic Sustainable Urban Mobility Plan to improve pedestrian safety satisfying the mobility needs of people and businesses in cities as well as their surroundings for a better quality of life. However, the city's road safety issues are still present and the status quo in the number of pedestrian injuries along with a very low perceived safety can seriously lead to fewer people moving by foot (European Commission, 2019c).

The severity of vehicle-pedestrian crashes confirmed that further actions to improve pedestrian safety are strongly needed. Road traffic crashes may be an everyday occurrence but they are both predictable and preventable as they take place due to a series of events influenced by several factors partly stochastic (casual) and partly deterministic (which can be identified). Nevertheless, in-depth pedestrian studies are required as risk factors associated with pedestrian-related crashes on transportation networks are usually different than for motor vehicles.

In so doing, identifying the most eligible method is a priority task. Furthermore, leaving aside the potential pros and cons of each method, data used to carry out crash severity analysis refer to crash data that are subject to extremely imbalanced distributions. The issue is related to the presence of one (or more than one) minority class. Thus, the minority classes (severe and fatal crashes) are likely to be masked by the majority classes (slight injuries in this case). The result is that, without taking the issue of imbalanced data into account, a classifier tends to predict the prevailing class more accurately than its minority counterparts. However, the minority classes are of significant value. Most learning methods are designed to identify the classification rule that best suits the data, according to a criterion of global accuracy, to minimize the global error to which the minority class contributes little (Ganganwar, 2012). The ineffectiveness of these algorithms in predicting the minority class, in presence of imbalanced data, is demonstrated by several studies. Both parametric and non-parametric methods are affected by the imbalanced distribution of the response classes (King 2001; Menardi & Torelli, 2012; Ndour et al., 2012). The problem of imbalanced distributions has been investigated in this research.

Crash severity analyses are generally carried out using crash data containing information recorded by police at the moment of the crash. Each crash should be described as more accurate as possible without missing values and avoiding inconsistencies. However, the real crash database contains just the information related to roadway, environment and people involved in the crash (age and gender) whereas no information about the driver behaviour as well as the pedestrian behaviour is generally provided. Thus, the analysts can focus on infrastructures aspects but they will never know what happened the instant immediately before the crash and they will ignore if a crash could have been a consequence of a behavioural factor. Moreover, the database structures among different countries considerably differ from each other as well as the information they stored.

Some databases, such as the British (provided by the UK government), the Spanish (provided by the DGT office), and the Swedish (provided by STRADA) crash databases are divided into more sections in which the information related to crash circumstances, vehicles involved and casualties are separately collected by the police. Such databases have a relaxed structure able to preserve all the data. Then, the three sub-datasets can be joined together by a crash code, which is unique for each crash. On the other hand, other datasets, such as the Italian one (provided by Istat) have a stricter structure. For instance, crash circumstances, vehicle info, and casualties' personal details are collected in the same datasheet. This represents a limitation in the information that can be reported for each crash. An example regards the maximum number of involved vehicles that can be collected. The Italian crash report form allows to record at most three vehicles. If more than three vehicles resulted involved in a crash, vehicles' information from the fourth vehicle onwards cannot be reported in the data. However, the greatest difference regards the crash severity variable. Many databases collect information related to crash severity on at least three different levels of severity. The British and the Spanish databases provide crash severity information as slight injury, serious injury, and fatal crashes collected by police at the scene of the crash. The Swedish databases, instead, collect crash severity info based on police and medical reports. The Swedish database is trying to introduce the Maximum Abbreviated Injury Scale (MAIS 3+) with the aim of standardising the definition of a serious injury in line with the EU recommendations. The Italian national database still continues to provide crash severity on two levels: fatal crash or injury, without mentioning the severity entity sustained by the people involved in the crash. However, this is too far from the European recommendation of providing crash severity at least on three different levels of information and does not allow researchers to focus their study also on serious crashes.

In this research, three national databases (Great Britain, Sweden, and Italy) were used in order to evaluate how the information reported in different data can impact the results provided by the model performances and results.

1.2 Main objectives

In light of the above, the main aim of this research is to investigate factors associated with pedestrian crashes by developing a profound analysis considering the coexistence of pedestrian, driver, vehicle, and environmental factors which may have caused an increase in pedestrian fatalities and severe injuries in recent years. Improving understanding of contributory factors can also assist in the selection of appropriate countermeasures for addressing pedestrian crash severity issues. Consequently, their identification and implementation require a thorough and accurate analysis of the factors that may mitigate or exacerbate the degree of injury sustained by pedestrians once a crash occurs.

Hence, the identification of the most suitable methodological approach to perform pedestrian crash severity analyses is crucial, also considering the critical issues related to the quality of data and the impact of the possible unavailable information.

The study of the literature highlighted the presence of two main groups of methods usually implemented in crash severity analyses. The two groups consist of the econometric models on one side and the machine learning tools on the other side. The research used data related to pedestrian crashes collected in different databases of Great Britain, Sweden and Italy. Thus, this research is also aimed at investigating and comparing the results obtained by both groups of models with respect to the same dependent variable (the crash severity in vehicle-pedestrian crashes).

In this research, the econometric models represented the starting point of the study. Their widespread use, so far, demonstrates the ability of these models in identifying patterns that occurrence may increase the severity of pedestrian crashes. Furthermore, the common expert wisdom has that the results obtained by the econometric models are considered reliable and accurate. Hence, this research aims at developing machine learning tools, which fall under the umbrella of artificial intelligence, and comparing econometric models and machine learning methods from a dual perspective: by their capability in identifying significant explanatory variables affecting crash severity (qualitative evaluation) and their model performances in predicting pedestrian crash severity by using performance metrics (quantitative evaluation). Then, the multi-model comparisons may guide future researchers in the choice of the best model to perform crash severity analyses.

The third aim of this research is to understand how to account for imbalanced data when performing the analysis in order to improve the model performance in terms of correct classification of the most serious pedestrian crashes.

Another aim of the research is also to understand how the ability of a model applied to different databases changes (different in structure, sample sizes, geographical conditions, and mainly, different in the way the variable crash severity is collected). So, for a certain method, the performances of that method will be also compared according to the results obtained for the different databases.

Then, both the results of traditional statistical methods and advanced data mining statistical techniques are investigated to find out relationships among crash data and to identify set of patterns that occurred together in a crash. Detected the interdependences between crash characteristics and crash pedestrian involvement, the research further aims to provide insights on possible improvements which can significantly enhance pedestrian safety and contribute to the vision zero-deaths on roads by 2050.

Finally, analysed the three different national databases, recommendations on how the Italian crash report form can be improved are provided.

Table 1 – Main objectives and research questions.

Objective	Research question
1. Investigation of the issue of imbalanced distributions of crash severity levels	How can imbalanced distributions be handled?
2. Comparison of the models in predicting pedestrian crash severity by using performance metrics (quantitative evaluation)	Is there a method prevailing over the others?
3. Understanding how the ability of a model applied to different databases changes (different in structure, sample size, geographical condition)	How does each model perform when applied in different context? Does it confirm its performances or not?
4. Comparison of the models by their capability in identifying significant explanatory variables affecting crash severity (qualitative evaluation)	Is there a method providing in-depth knowledge and non-trivial relations among data?
5. Identification of the interdependences between crash characteristics and fatal pedestrian crashes	Which factors mostly contribute to fatal pedestrian crashes?
6. Proposal for improvement of the Italian crash report form	What information can be further collected by the authorities?

The main novelties and objectives of this research were resumed below. The remaining of this work was organized as descripted in paragraph 1.4. In addition, a sequence of specific steps was provided in order to give a whole framework on which were the goals to be achieved step by step (Table 2).

1.3 Research novelties

The eight-fold novelties and main contributions of this PhD research can be divided into methodological and engineering points.

As for the methodological contributions:

1) The research comprehensively discusses the concern about pedestrian safety and the existing models commonly performed in crash prediction studies.

2) The literature review of this research describes the current methodological framework more clearly. Thus, resuming tables are provided from different perspectives:

- Prior research on pedestrian crash severity which used econometric models
- Prior research on pedestrian crash severity which used machine learning algorithms
- For the prior research, summary information about the study period and the sample size used in the analyses is provided
- Main significant results of the previous research are reported in the literature review and they were divided according to the roadway, environment, vehicle, crash, driver, and pedestrian identified factors contributing to pedestrian severity.

3) The work points out some significant research questions which are still unanswered according to the recent literature:

- Which is the most suitable method to perform crash severity analyses focused on pedestrians?
- Machine learning algorithms have been used for crash severity prediction in lieu of the more traditional econometric models. Are these tools effective in providing meaningful insights on potential factors crash contributing to pedestrian severity?
- How do the performances of econometric models and machine learning algorithms change in presence of different data-sample sizes?
- Is there a method able to uncover contributory factors with the highest possible prediction accuracy?
- Do econometric models and machine learning algorithms provide similar results or there are significant dissimilarities?
- How does each model’s performances vary when developed using different crash databases?

4) There is not a comprehensive comparison between econometric models and machine learning tools that covers both prediction accuracy and contributory factors, two equally important subjects in pedestrian crash severity studies. This research addresses this gap by examining the key differences

between econometric models and machine learning algorithms. The research provides an in-depth multi-model comparative analysis by developing four econometric models and five machine learning tools using three different databases (British, Swedish and Italian data at national level).

5) The research investigates the issue related to imbalanced data and applied a weighted approach to take the issue into account in models' development.

6) It indicates the advantages of using a model, or a combination of models, which may provide support for the choice of the appropriate prediction method for crash severity analyses with unbalanced data.

7) From the engineering perspective, this research explores the explanatory variables identified by the methods as affecting pedestrian crash severity.

8) On the basis of the potential contributory factors to fatal and serious pedestrian crashes, safety countermeasures to mitigate and minimize pedestrian crash severity are proposed.

The potential impact the research will have on the transportation field consists of a significant advance over past works done in crash severity prediction by providing a comparison of methods commonly used to analyse crash severity. Even though the models have already been applied to model pedestrian injury severity, a comparative analysis to assess the predictive power of such modeling techniques is limited. The comparison will be provided on a double scale: each model has been applied to different databases so firstly their response was analysed, meant to understand the model classification ability, and then, on the same dataset, the different models were compared in terms of quantitative and qualitative performances.

1.4 Thesis organization

In this section, a brief and concise description of the structure of this thesis work was provided.

This thesis elaboration is organised as follows:

Chapter 1 presents the objectives to be fulfilled in this thesis work and the work organization.

Chapter 2 includes an overview of global and European road safety issues. Then, the literature review focused on pedestrian statistics and concerns over time. A paragraph of the chapter is characterised by a brief description of the European road safety plan and resume the efforts that EU Commission is singling out. Few comments on crash databases' similarities and differences were also provided. The latter also includes the definition of crash severity and a resume of the main contributory factors identified by previous research on pedestrian severity as well as the methods adopted and the sample size.

Chapter 3 focuses on the methodological approach describing the 9 methods that will be performed successively. The second part is characterised by the weighted approach used to take the issue of imbalanced data into account and the measures of performance used to evaluate the reliability of the models.

Chapter 4 presents the study context. This research applied both econometric models and data mining tools on three National databases. Thus, this chapter provides an overview of each database used in this research and insights on how data are collected and stored in Great Britain, Sweden, and Italy. Descriptive statistics are provided as well in order to resume the main variables of interest for each database.

Chapter 5 shows the results of modelling pedestrian crash severity analysis using both econometric models and machine learning tools. Results are provided for each country, in different sections of the chapter.

Chapter 6 provides a discussion of the main results and the considerations derived from this thesis work. Possible safety improvements strategies focused on pedestrians are provided.

Chapter 7 provides a summary of the dissertation phases and draw the main conclusions of the research.

Appendixes Great Britain, Sweden, Italy contains all results the models provided for Great Britain, Sweden, and Italy which are not included in chapter 5.

The specific step by step objectives meant to be achieved in this work are reported as follow in the Table 2.

Table 2 – Doctoral thesis framework.

1st step	Analysis of Road Safety worldwide and crash fatality trends overtime focusing on pedestrian crashes and the most severe consequences. Is the trend worrying?
2nd step	Description of crash databases: how they are organised, how crash severity is collected, which serious injury definition is adopted. In which way could the Italian national database be improved?
3rd step	In-depth analysis of the most popular methods used to assess pedestrian crash severity over time. Which are the most used models over time? On which sample sizes had these models been implemented?
4th step	Resume of the main significant results of prior research on potential factors contributing to pedestrian crash severity and their impact on pedestrian severity. What has been already found by prior research?
5th step	Fatal and serious crashes represent a small share of the total crashes. Are the traditional approaches able to capture the minority class or they focus most on the majority group? How can imbalanced distributions be handled?
6th step	In-depth analysis of the most popular performance metrics used to assess method classification ability in previous research on pedestrian crash severity over time. Which performance metrics could be used to compare econometric models and machine learning tools?
7th step	Implementation of both econometric methods (traditional methods to assess pedestrian crash severity) and machine learning algorithms (representative of the new methodological frontier) on 3 National databases: Great Britain, Sweden, and Italy. How do the models respond?

8th step	Quantitative comparison among the implemented methods in terms of accuracy and reliability of the results (for each database separately). Is there a method prevailing over the others?
9th step	Exploration of how classification performances for each model may vary depending on: different sample sizes and a different number of independent variables.
10th step	Qualitative considerations on the potential contributory factors identified by the implemented methods (for each database separately). Is there a method providing in-depth knowledge and non-trivial relations among data?
11th step	Quantitative comparison of the performance of the same model developed in the three different National crash databases. Are the models, both econometric and machine learning tools, sensible to the geographical context?
12th step	Identification of the most appropriate model (or a combination of models) to conduct pedestrian crash severity analysis. Which models can be used to explore crash severity factors?
13th step	Insights on the possible factors contributing to the most severe pedestrian crashes more often than they would if they were independent of each other and the possible safety improvement strategies focused on pedestrians. Do some factors contribute to pedestrian crashes in all databases investigated?

CHAPTER II ~ LITERATURE REVIEW

In this chapter, an overview of Road Safety worldwide statistics involving pedestrian victims, road safety plans and crash databases was provided.

At the beginning, it was analysed the number of crash deaths that occurred each year as a consequence of a road crash for different countries. Then, statistics focusing on road deaths collected in Europe were provided followed by vulnerable road users and pedestrian statistics. Furthermore, an overview of road safety plans was provided to summarize the efforts carry out worldwide to reduce crash severity. Best practices and actions were mentioned as well. Finally, a critical overview of international crash databases was presented followed by the most popular statistical methods used in prior research carried on pedestrian crashes and then, their main contributions were summarized.

2.1 Road safety statistics

Worldwide, road crashes constitute a real concern and a serious public health problem for each country (Table 3). Despite progress has been achieved in important areas such as legislation, vehicle standards and improving access to post-crash care, this progress has not, however, occurred rapidly enough to compensate for the rising population and rapid demand for transport taking place in many parts of the world. Every year the lives of more than 1.25 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability because of their injury (WHO, 2018). Children and older people as well as pedestrians, cyclists, and motorcyclists are considered the most vulnerable of road users.

2.1.1 International road safety statistics

The number of deaths due to road crashes was reported for Europe, the United States, Russia, Japan, and Australia since 2000. The U.S. and Europe recorded the major number of deaths due to road crashes, exhibiting a quite constant trend. After an initial decrease in crash deaths recorded in the U.S. between the ten-year period 2000-2010, the trend remains unchanged with very slightly fluctuation. In Europe, the trend is unfortunately consistent with the American one with very intangible changes (Table 3).

Table 3 – Road deaths for EU, USA, Russia, Japan and Australia overtime (Sources: EUROSTAT 2021 & IRTAD Safety Annual Report 2020).

	2000	2010	2015	2016	2017	2018	2019
EU27	53,502	29,611	24,360	23,814	23,393	23,331	22,763
U.S.	41,945	32,999	35,092	27,875	37,373	36,560	36,120
Russia	-	26,567	23,114	20,308	19,088	18,214	16,981
Japan	10,410	5,828	4,867	4,698	4,431	4,166	3,920
Australia	1,817	1,351	1,205	1,294	1,225	1,136	1,189

For the same countries, the number of road deaths was evaluated as a function of population. Russia and U.S. recorded each year a total number of road crash victims quite doubled than Europe and almost three times more than Japan and Australia. However, even if Australia was characterized by the lowest death number on roads, statistics reveal that the rate of road deaths evaluated per 100,000 inhabitants (equal to 4.7 in 2019) was comparable to the European rate (equal to 5.1 in 2019) as Australia is less populated than Europe and accounting for 25,364,307 inhabitants compared with 447,512,041 Europeans in the same year (Table 4).

Table 4 – Road deaths per 100,000 inhabitants for EU, USA, Russia, Japan and Australia (Sources: EUROSTAT 2021 & IRTAD Safety Annual Report 2020).

	2000	2010	2015	2016	2017	2018	2019
EU27	-	6.6	5.2	5.8	5.7	5.2	5.1
U.S.	14.9	10.7	10.9	8.6	11.5	11.2	11.0
Russia	20.2	18.6	16.0	14.1	13.2	12.6	11.8
Japan	8.2	4.6	3.8	3.7	3.5	3.3	3.1
Australia	9.5	6.1	5.1	5.3	5.0	4.5	4.7

Moreover, all countries exhibited the same trend: a slight, almost flat reduction in road fatalities especially during the last decade.

2.1.2 European road safety statistics

With the aim of improving mobility and enhancing road safety, the European Commission has embarked Road Safety Actions since 1998. Moreover, according to EU27 statistics, the total deaths on road reached 51,351 units only in 2001 (Figure 1). Thus, it was since 2001 that the European Commission started promoting decades of actions to reduce the serious number of crashes involved in people reclaiming the commitment of each EU Member State.

The first European transport decennial policy was set out in 2001 and it was called “time to decide” with the principal aim of halving the number of road deaths in the period 2001-2010. The EU27 target was not achieved and 29,611 fatalities (out of the desired 26,676 needed to meet the goal) were recorded in Europe (Figure 1). Thus, in 2010, with the goal to “stabilize and then reduce” the predicted increase in road traffic fatalities and in agreement with the UN General Assembly, the European Commission proposed to continue with the same target of halving the overall number of road fatalities by 2020 using 2010 as a baseline. The new decade was declared “Decade of Action for Road Safety” and the main goals were reported in the plan “Towards a European road safety area: policy orientations on road safety 2011-2020”. The need for a coherent holistic and integrated approach required taking into account synergies with other policy goals. Thus, road safety policies at local, national, European or international levels were integrated with relevant objectives of other public policies and vice versa.

According to the most recent and preliminary statistics provided by the official EU website (https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1767), an estimated 18,800 people were killed in a road crash in 2020, an unprecedented annual fall of 17% in 2019. 4,000 fewer people lost their lives on EU27 roads in 2020 compared to the preceding year. However, the reduction was far from the desired progress (Table 5) and the plan 2021-2020 failed its target of halving road fatalities in ten years with the number of road deaths dropping by only 32% over the previous decade between 2010 and 2020 (Figure 1).

Since the first EU target for reducing the number of road deaths was introduced in 2001, Lithuania achieved the strongest reduction in road fatalities (less 76% assessed between 2001 and 2019) followed by Estonia and Lithuania with 74% fewer fatalities in 2019 compared with 2001. Latvija, Estonia, Lithuania, Luxembourg, Spain, Ireland, Greece, Slovenia, Sweden, France, Portugal, Slovak Republic, Austria, Belgium, Germany, Italy, Croatia, Denmark, Czechia, Hungary, and Finland were the 21 countries whose progress have recorded at least 50% fewer road deaths. Since 2010, the greatest progress has been recorded in Greece, and confirmed in Latvija and Lithuania. However, in other countries, such as Romania and France, the progress has flattened out whereas the decrease in road fatalities has reversed its trend in Malta and Netherlands where an increase in fatalities has been recorded in 2019 compared to 2010. Nevertheless, it is noteworthy to say that the number of road traffic fatalities in the various countries and regions depends on both structural differences (size of the country/region; composition, density and quality of the road network, characteristics of the population) and socio-economic differences (characteristics of the vehicle stock, transit and tourist traffic, behavioural aspects,…).

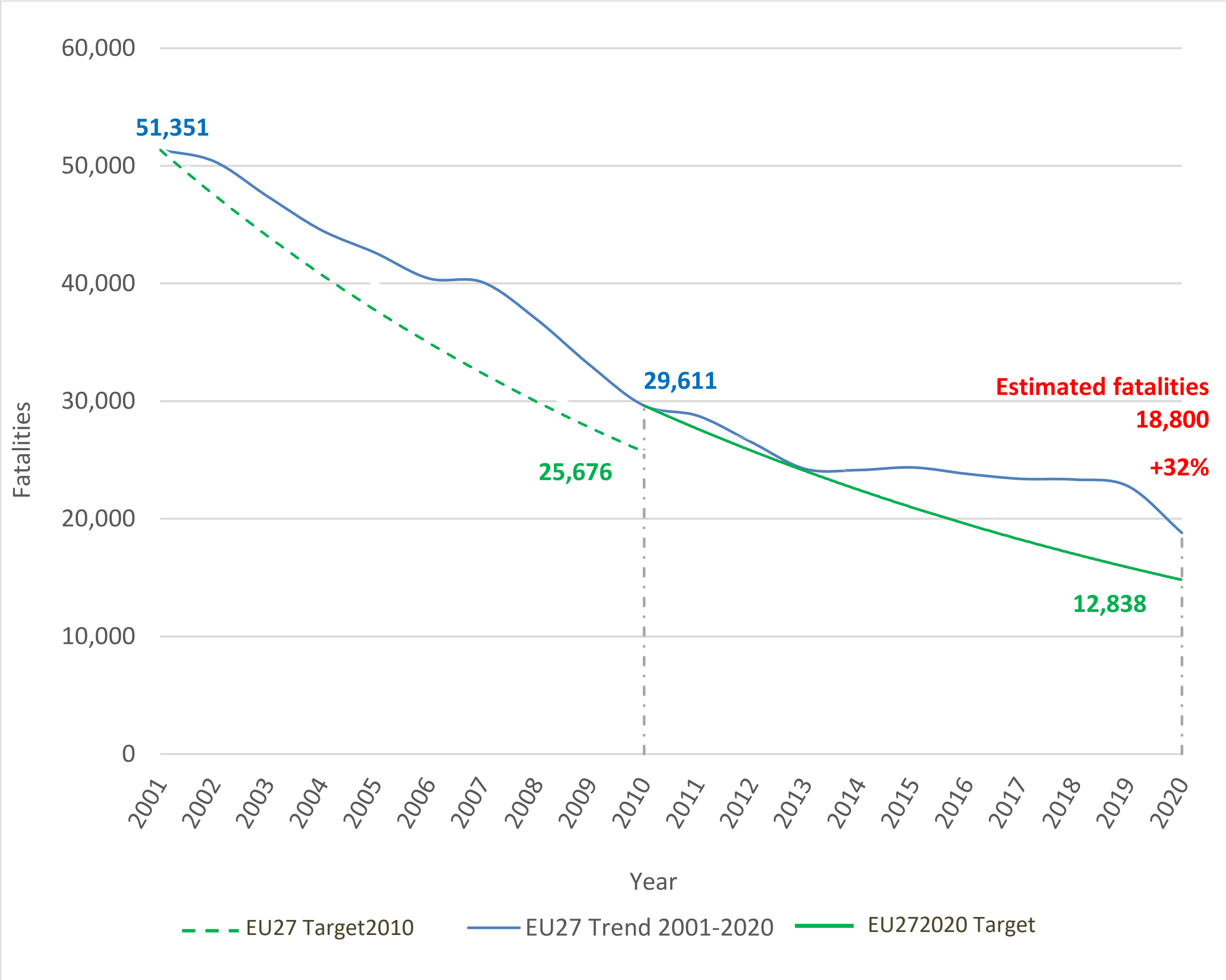


Figure 1 – EU27 Road Safety Targets: actual and desired progress (Source: EUROSTAT, 2021).

As follow, road fatalities and relative changes between 2001 and 2019 were provided. Data were retrieved from the official EU website (EU statistical pocketbook 2021, https://ec.europa.eu/transport/facts-fundings/statistics/pocketbook-2021).

Table 5 – Road deaths and and their relative long-term change between 2001 and 2019 (Sources: CARE & EUROSTAT, 2021).

Country	2001	2005	2010	2015	2016	2017	2018	2019	2001-2019 %	2010-2019 %
Austria	958	768	552	479	432	413	409	416	-56.6	-24.6
Belgium	1,486	1,089	840	732	637	620	604	646	-56.5	-23.1
Bulgaria	1,011	957	776	708	708	682	610	628	-37.9	-19.1
Croatia	647	597	426	348	307	331	317	297	-54.1	-30.3
Cyprus	98	102	60	57	46	53	49	52	-46.9	-13.3
Czechia	1,333	1,286	802	734	611	577	656	618	-53.6	-22.9
Denmark	431	331	255	178	211	183	171	199	-53.8	-22.0
Estonia	199	170	79	67	71	48	67	52	-73.9	-34.2
Finland	433	379	272	266	258	223	239	211	-51.3	-22.4
France	8,162	5,318	3,992	3,461	3,477	3,456	3,246	3,244	-60.3	-18.7
Germany	6,977	5,361	3,648	3,459	3,206	3,177	3,275	3,046	-56.3	-16.5
Greece	1,880	1,658	1,258	793	824	739	700	688	-63.4	-45.3
Netherlands	993	750	537	531	533	613	598	586	-41.0	9.1
Hungary	1,239	1,278	740	644	607	624	633	602	-51.4	-18.6
Ireland	412	400	212	162	186	158	138	140	-66.0	-34.0
Italy	7,096	5,818	4,114	3,428	3,283	3,378	3,334	3,173	-55.3	-22.9
Latvija	558	442	218	188	158	136	148	132	-76.3	-39.4
Lithuania	706	773	299	242	192	192	173	186	-73.7	-37.8
Luxembourg	70	47	32	36	32	25	36	22	-68.6	-31.3
Malta	16	17	13	11	23	19	18	16	-	23.1
Poland	5,534	5,444	3,908	2,938	3,026	2,831	2,862	2,909	-47.4	-25.6
Portugal	1,670	1,247	937	593	563	624	700	688	-58.8	-26.6
Romania	2,450	2,629	2,377	1,893	1,915	1,951	1,867	1,864	-23.9	-21.6
Slovak Republic	625	606	353	310	275	276	260	270	-56.8	-23.5
Slovenia	278	258	138	120	130	104	91	102	-63.3	-26.1
Spain	5,517	4,442	2,479	1,689	1,810	1,846	1,806	1,755	-68.2	-29.2
Sweden	583	440	266	259	270	253	324	221	-62.1	-16.9
EU 27	51,351	42,607	29,611	24,360	23,814	23,393	23,332	22,763	**-55.7**	**-23.1**

To better understand the real trend recorded by each Member State over years, the values reported in the previous table were also graphically provided. Below, a glance at trends is provided in terms of how many lives were lost due to road traffic crashes in the European Union over years. In white were represented the road deaths related to 2001, while in red and in blue were represented the evolutions of the road deaths number, respectively concerning 2010 and 2019. Road fatalities remain still high in France, Italy, Germany, Spain and Poland.

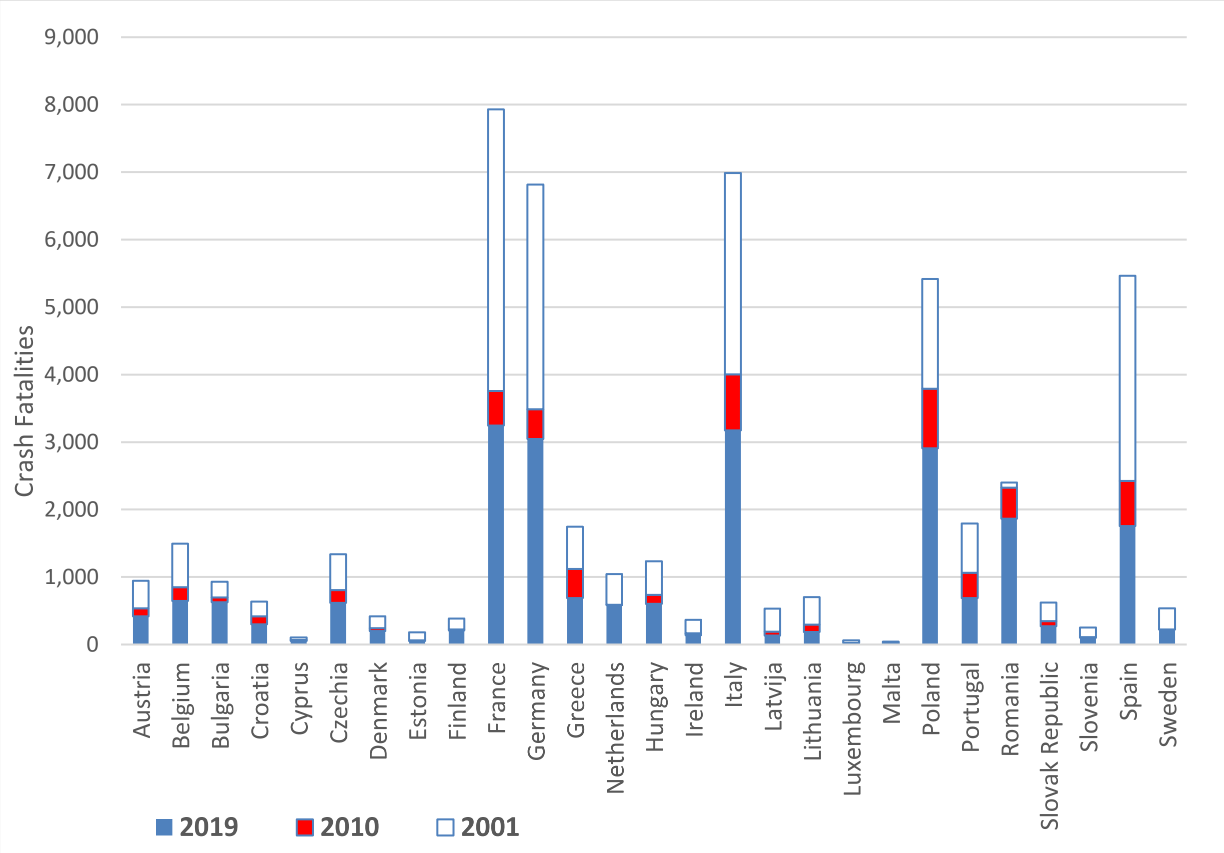


Figure 2 – Fatal Crashes in Europe (Sources: CARE & EUROSTAT, 2021).

Furthermore, focusing the attention on death variation percentage recorded in just one year from 2018 to 2019, what emerged is that even though the progress rate slowed down (In some countries, indeed, in 2019, the number of crash fatalities recorded is higher than the year before. This increase was observed especially for Denmark and Slovenia and a global reduction trend can be observed in road fatalities over the last decade, further efforts, as well as considerable improvements, are needed by all actors to improve road safety. In some countries, indeed, in 2019, the number of crash fatalities recorded is higher than the year before. This increase was observed especially for Denmark and Slovenia.

Table 6 – Road deaths and their relative short-term change between 2018 and 2019 in EU27 (Sources: CARE & EUROSTAT, 2021).

Country	**2018**	**2019**	**2018-2019 %**
Austria	409	416	1.7
Belgium	604	646	7.0
Bulgaria	610	628	3.0
Croatia	317	297	-6.3
Cyprus	49	52	6.1
Czechia	656	618	-5.8
Denmark	171	199	16.4
Estonia	67	52	-22.4
Finland	239	211	-11.7
France	3,246	3,244	-0.1
Germany	3,275	3,046	-7.0
Greece	700	688	-1.7
Hungary	633	602	-4.9
Ireland	138	140	1.4
Italy	3,334	3,173	-4.8
Latvija	148	132	-10.8
Lithuania	173	186	7.5
Luxembourg	36	22	-38.9
Malta	18	16	-11.1
Netherlands	598	586	-2.0
Poland	2,862	2,909	1.6
Portugal	700	688	-1.7
Romania	1,867	1,864	-0.2
Slovak Republic	260	270	3.8
Slovenia	91	102	12.1
Spain	1,806	1,755	-2.8
Sweden	324	221	-31.8
EU 27	23,331	22,763	**-2.4**

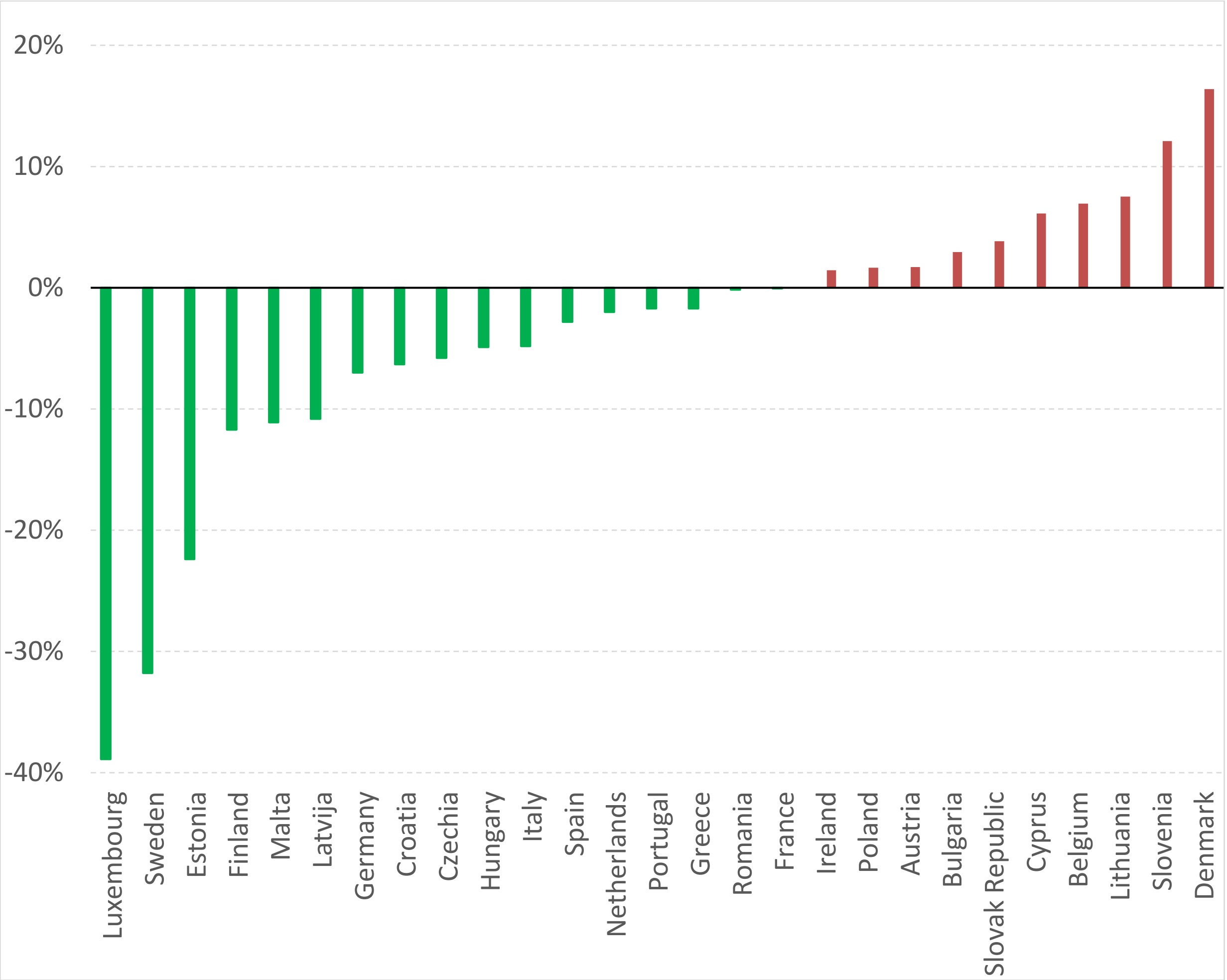


Figure 3 – Change in road deaths in EU27 2018-2019 (Sources: CARE & EUROSTAT, 2021).

Figure 3 shows that not enough had been done in road safety from 2018 to 2019. Deaths on road have increased in ten European nations in 2019 compared with the year before. Road fatalities mainly increased in Denmark (by 16%) and in Slovenia (by 12%) showing that more efforts need to be continued and further strengthened in way to achieve the new ambitious predefined target of halving road deaths and serious injury by 2030. In two nations (France and Romania) the change in the number of road fatalities was almost equal to zero. The other nations reached a decrease in the number of road traffic deaths, however, just a handful of nations achieved a reduction greater than 10%.

Finally, Table 7 provides different ranks for European nations established on the basis of road fatalities per million inhabitants, per ten billion pkm, and per million passenger cars and the average values for EU27 (without considering the United Kingdom). Note that with "fatalities" European Commission means all fatalities on the road: car drivers and passengers, bus and coach occupants, powered two-wheelers' riders and passengers, cyclists, pedestrians, commercial vehicle drivers. Pkm is an indicator of traffic volume (in the absence of consistent vehicle-kilometre data); passenger-kilometres of cars plus (mostly estimated) passenger-kilometres of motorised two-wheelers. The number of inhabitants is the average population in 2019 provided by Eurostat, and with passenger cars, the European Commission means the average stock of vehicles for 2018 and 2019 (European Commission, 2021).

Table 7 – Road fatalities in EU in 2019 (Sources: CARE & EUROSTAT, 2021).

Per million inhabitants		**Per 10 billion pkm**		**Per million passenger cars**	
Country	**Road fatalities**	**Country**	**Road fatalities**	**Country**	**Road fatalities**
Sweden	22	Sweden	19	Sweden	45
Ireland	28	Ireland	24	Luxembourg	52
Malta	32	Luxembourg	27	Malta	53
Netherlands	34	Finland	31	Finland	60
Denmark	34	Denmark	32	Germany	64
Luxembourg	35	Germany	33	Ireland	65
Germany	37	Estonia	36	Estonia	67
Spain	37	Slovenia	36	Netherlands	68
Finland	38	Netherlands	39	Spain	72
Estonia	39	France	41	Denmark	76
Austria	47	Italy	41	Italy	81
France	48	Spain	50	Austria	83
Slovenia	49	**EU27**	**51**	France	85
Slovak Republic	50	Austria	51	Slovenia	88
EU27	**51**	Malta	56	Cyprus	93
Italy	53	Lithuania	59	**EU27**	**93**
Belgium	56	Belgium	59	Czechia	106
Czechia	58	Greece	62	Belgium	110
Cyprus	59	Portugal	69	Slovak Republic	115
Hungary	62	Czechia	72	Poland	122
Greece	64	Cyprus	73	Lithuania	127
Lithuania	67	Latvija	82	Portugal	128
Portugal	67	Hungary	88	Greece	129
Latvija	69	Slovak Republic	90	Hungary	162
Croatia	73	Bulgaria	105	Croatia	175
Poland	77	Croatia	116	Latvija	184
Bulgaria	90	Poland	117	Bulgaria	224
Romania	96	Romania	162	Romania	279

2.1.3 Vulnerable road user statistics

The number of road fatalities recorded each year is unacceptably high. Road crash death for all age groups even surpassed HIV/AIDS, tuberculosis and diarrhoeal diseases (WHO, 2018). As matter of fact, currently, road traffic injury is the leading cause of death for children and young adults aged between 5–29 years. What is more, the worldwide burden of road traffic injuries and deaths is disproportionately borne by vulnerable road users (VRU) which include children, older people, pedestrians, cyclists, and motorcyclists (Table 8). Each year, indeed, almost 50% of all road fatalities is covered by VRU fatalities.

The increasing number of crashes involving VRU and their crash fatality statistics represent the most serious challenge for the new decade of action for road safety.

Table 8 – Annual number of VRU fatalities, and their share in the total number of fatalities in the EU27 in 2010-2019 (Sources: CARE & EUROSTAT, 2021).

Year	EU27 fatalities	VRU fatalities	VRU as % of total
2010	29,611	13,179	44.5
2011	28,725	13,174	45.9
2012	26,440	12,118	45.8
2013	24,227	11,481	47.4
2014	24,141	11,482	47.6
2015	24,360	11,247	46.2
2016	23,814	10,935	45.9
2017	23,393	10,892	46.6
2018	23,332	10,959	47.0
2019	22,763	11,433	48.2

In Table 9 statistics related to PTW riders, cyclists, and pedestrians are provided as well as their share in the total fatalities occurred in the 27 countries of European Union. Among vulnerable road users, pedestrians are the most susceptible to road potential risks and they form the second largest group of road casualties after car occupants. The annual number of pedestrian fatalities, and their share in the total number of fatalities in the EU27 from 2010 to 2018 is reported.

Table 9 – Annual number of PTW rider, cyclist, and pedestrian fatalities, and their share in the total number of fatalities in the EU27 in 2010-2019 (Sources: CARE & EUROSTAT, 2021).

Year	EU27 fatalities	VRU fatalities					
		PTW rider		Cyclist		Pedestrian	
		N	%	N	%	N	%
2010	29,611	5,242	17.7	1,985	6.7	5,952	20.1
2011	28,725	5,201	18.1	1,989	6.9	5,984	20.8
2012	26,440	4,615	17.5	2,075	7.8	5,428	20.5
2013	24,227	4,252	17.6	1,921	7.9	5,308	21.9
2014	24,141	4,192	17.4	2,043	8.5	5,247	21.7
2015	24,360	4,274	17.5	1,975	8.1	4,998	20.5
2016	23,814	3,969	16.7	2,000	8.4	4,966	20.9
2017	23,393	4,101	17.5	1,921	8.2	4,870	20.8
2018	23,332	4,190	18.0	2,006	8.6	4,763	20.4
2019	22,763	4,162	18.3	2,147	9.4	4,652	20.4

The role of pedestrians in the traffic and transport system is essential. Several studies show that the proportion of trips on foot varies between 8% and 27% (OECD, 2010) and the share of walking is even higher for short trips under 5 km (Wittink, 2001). However, the number of pedestrian crashes recorded each year worldwide, as well as the severity of pedestrians' outcomes, is evidence that pedestrian safety is a real global concern at social, political, economic and engineering levels. As matter of fact, every journey includes a walking component and it is also a fundamental human activity. Thus, people may have preferences when it comes to the transportation mode, but at some point, everyone is a pedestrian and everybody moves on foot even just for relatively short distances. However, walking is often overlooked or given low priority when planning and designing transport networks (Austroads, 2020). The result is that pedestrians are the most vulnerable road users and are at risk of the most severe consequences when involved in traffic crashes (WHO, 2018). The evidence is provided in Table 9: each year, the number of pedestrian fatalities alone accounts at least for one-fourth of the total road victims.

In 2019, pedestrian deaths were reduced by 22% on average among all EU countries compared to 2010 (Table 10). In all 27 countries, there was actually a decrease in pedestrian fatalities albeit the progress stagnated in Cyprus, France, and Estonia.

Table 10 – Pedestrian fatalities and their relative long-term change between 2010 and 2019 in EU27 (Sources: CARE & EUROSTAT, 2021).

Country	2010	2019	2010-2019 %
Austria	98	69	-29.6
Belgium	108	92	-14.8
Bulgaria	174	154	-11.5
Croatia	105	61	-41.9
Cyprus	13	13	0.0
Czechia	168	111	-33.9
Denmark	44	30	-31.8
Estonia	14	13	-7.1
Finland	35	15	-57.1
France	485	476	-1.9
Germany	476	421	-11.6
Greece	179	145	-19.0
Hungary	192	144	-25.0
Ireland	44	35	-20.5
Italy	621	534	-14.0
Latvija	79	40	-49.4
Lithuania	96	81	-15.6
Luxembourg	1	2	100.0
Malta	5	2	-60.0
Netherlands	62	49	-21.0
Poland	1,236	793	-35.8
Portugal	195	140	-28.2
Romania	868	729	-16.0
Slovak Republic	126	80	-36.5
Slovenia	26	15	-42.3
Spain	471	381	-19.2
Sweden	31	27	-12.9
EU27	5,952	4,652	-21.8

Table 11 – Number of pedestrian fatalities in the total number of fatalities, per country in the EU27 in 2019 (Sources: CARE & EUROSTAT, 2021).

Country	**Pedestrian fatalities**	**All fatalities**	**Proportion of pedestrian fatalities %**
Austria	69	416	16.6
Belgium	92	646	14.2
Bulgaria	154	628	24.5
Croatia	61	297	20.5
Cyprus	13	52	25.0
Czechia	111	618	18.0
Denmark	30	199	15.1
Estonia	13	52	25.0
Finland	15	211	7.1
France	476	3,237	14.7
Germany	421	3,046	13.8
Greece	145	688	21.1
Hungary	144	602	23.9
Ireland	35	182	19.2
Italy	534	3,173	16.8
Latvija	40	132	30.3
Lithuania	81	242	33.5
Luxembourg	2	22	9.1
Malta	2	18	11.1
Netherlands	49	586	8.4
Poland	793	2,909	27.3
Portugal	140	688	20.3
Romania	729	1,864	39.1
Slovak Republic	80	270	29.6
Slovenia	15	102	14.7
Spain	381	1,755	21.7
Sweden	27	221	12.2
EU27	4,652	22,763	**20.4**

At the end of 2019, the pedestrian road safety progress was as reported in (Table 11). Yet 4,652 pedestrians were killed only in 2019 representing 20.4% of all road victims (Table 11). Furthermore, their share in the total number of fatalities observed in 2019 was actually the same observed ten years before. Nevertheless, considerable disparities in pedestrian safety exist in Europe. Pedestrian road fatality rate varies by almost a factor of ten between countries (Table 12). Some countries, namely Romania and Lithuania, scored the highest pedestrian road fatality rate among EU countries with 37.5 and 29.0 deaths per million inhabitants, respectively. These rates are more than the triple the EU average of 10.4. Despite the positive developments in reducing the number of pedestrian deaths (Table 11), other countries, namely Bulgaria, Poland, and Latvija, still scored higher fatality rates in the EU27 (Table 12) equal to 22.0, 20.9, and 20.8 per million inhabitants in 2019, respectively. In particular, these nations are those occupied the lowest ranks when fatality rate was assessed considering all road fatalities.

Trying to reverse this trend, many nations have been considering actions to improve pedestrian safety in most national road safety action plans. Just to provide examples of this commitment, since 2008, France has been proposing the "code de la rue" approach (Decree 2008-754) which is a change in the system of traffic rules to increase awareness of and respect for the safety of the most vulnerable road users. In Belgium multiple cities and towns in Belgium are “pedestrianizing” specific areas by introducing or extending 30 km/h zones and traffic calming measures. In Greece, the national road safety strategy report 52 (National Technical University of Athens, 2011) proposed adaption of traffic light intervals to take into account the slower walking pace of elderly or people with disabilities.

Table 12 – Pedestrian fatalities per million inhabitants per nation in the EU27 in 2019 (Sources: CARE & EUROSTAT, 2021).

Country	**Pedestrian fatalities**	**Population (per million inhabitants)**	**Fatality rate**
Austria	69	8.86	7.8
Belgium	92	11.46	8.0
Bulgaria	154	7.00	22.0
Croatia	61	4.08	15.0
Cyprus	13	0.88	14.8
Czechia	111	10.65	10.4
Denmark	30	5.81	5.2
Estonia	13	1.32	9.8
Finland	15	5.52	2.7
France	476	67.18	7.1
Germany	421	83.02	5.1
Greece	145	10.72	13.5
Hungary	144	9.77	14.7
Ireland	35	4.90	7.1
Italy	534	59.82	8.9
Latvija	40	1.92	20.8
Lithuania	81	2.79	29.0
Luxembourg	2	0.61	3.3
Malta	2	0.49	4.1
Netherlands	49	17.28	2.8
Poland	793	37.97	20.9
Portugal	140	10.28	13.6
Romania	729	19.41	37.5
Slovak Republic	80	5.45	14.7
Slovenia	15	2.08	7.2
Spain	381	46.94	8.1
Sweden	27	10.23	2.6
EU27	4,652	446.45	10.4

2.2 Road safety plans

The problem of deaths and injuries as a result of road crashes is acknowledged to be a global phenomenon. In the Resolution adopted by the General Assembly on 31 August 2020 it was said that "*Road traffic deaths and injuries are also a social equity issue, as the poor and the vulnerable are most frequently also vulnerable road users, namely, pedestrians, cyclists, users of motorized two- and three-wheeled vehicles and passengers of unsafe public transport, who are disproportionately affected and exposed to risks and road traffic crashes, […], and the aim of road safety policies should be to guarantee protection to all users"*.

With the shared aim of reducing crash severity, the General Assembly proclaims the period 2021–2030 as the Second Decade of Action for Road Safety and calls on the Member States to continue action through 2030 on all the road safety-related targets of the Sustainable Development Goals. Indeed, among the 17 Sustainable Goals set in the 2030 Agenda (https://sdgs.un.org/goals), not only do the United Nations encourage actions to halve the number of global deaths and serious injuries from road traffic crashes by 2020 (goal 3.6), they also recognize the importance of providing access to safe, affordable, accessible and sustainable transport systems for all by 2030, improving road safety, with special attention to the needs of vulnerable users (goal 11.2). European Commission welcomed the launch of the Global Plan for the UN Decade of Action on Road Safety 2021-2030 the 28th of October, 2021 setting out how to achieve the target to reduce road traffic deaths and injuries by 50% by 2030.

With the new ten-year European Road Safety Policy Framework 2021-2030 – Next steps towards "Vision Zero", the EU has reaffirmed its ambitious long-term goal, to move close to zero deaths by 2050 (European Commission, 2020). By endorsing the Valletta Declaration on road safety of March 2017 in Council conclusions, EU transport ministers also, for the first time, set a target for reducing serious injuries, namely to halve the number of serious injuries in the EU by 2030 from a 2020 baseline. The plan has the long-term goal of zero deaths and serious injuries on roads by 2050. The achievement of the long term goal includes the achievement of the interim targets of recording 50% fewer deaths and serious injuries between 2020 and 2030. The success of the plan also relies on setting timed targets. At this aim, the European Commission in close cooperation with the Member States has established a set of intermediate outcome targets based on Key Performance Indicators (KPI) to monitor progress and directly linked to reducing deaths and injuries. Coordinated actions of all sectors and for all road users are further required to strengthen road safety efforts and spread good practice both inside the EU and internationally.

The road safety plan, guided by the "safe system approach" (European Commission, 2020) embraces the UN goals and invites the Member States to reconfirm their commitments and suggest main research should be oriented towards vulnerable road users. The safe system approach relies on 5 pillars: road safety management, safer roads and mobility, safer vehicles, safer road users, and rapid post-crash response which also consist in the set of themes to tackle the biggest road safety challenges.

The main targets of the five pillars are resumed below:

1) Ensuring safer roads and safer roadsides by matching road function, design, and layouts to accommodate human error in a way that crashes do not lead to serious consequences. In support of road safety, a systematic risk mapping and a safety rating provide useful tools to assess the status of the road network and to target investment. Recently, the EU has mandated the risk mapping and safety rating for roads belonging to the Trans-European Transport Network (TEN-T). KPI for infrastructure: Percentage of distance driven over roads with a safety rating above an agreed threshold. Indeed, the European Road Assessment Programme (EuroRAP), an international non-profit organisation of automobile clubs, road authorities and researchers, has carried out a systematic risk mapping and safety rating which result in ratings for roads between 1 and 5 stars. The approach also reflects a proactive assessment in addition to the more traditional reactive analysis of crash hot spots and provides a useful tool to assess the safety quality of the road network and to target investment.

2) Providing vehicles with innovations in technology equipment both to mitigate the crash severity and reduce the likelihood of crashes through passive elements such as safety belts and airbags, and active safety features, such as Advanced Emergency Braking, Intelligent Speed Assistance, Stability Control and Lane Departure Warning that may prevent crashes from happening altogether. Moreover, to ensure that users are protected through the lifetime of the vehicles, regular roadworthiness checks are scheduled. A KPI for vehicle safety considers the percentage of new passenger cars with a Euro NCAP safety rating equal or above a predefined threshold (e.g. 4-star) – to be specified further.

3) Preventing and mitigating fatalities and serious injuries due to road crashes embraces the third pillar. Member States will promote safer road use (speed, driving without alcohol and drugs, undistracted driving, safety belt and child restraint use, helmet use) by establishing stringent requirements for driver licensing, targeted education and professional drivers' training, supported by strong and sustained compliance and enforcement regimes with the pivotal role of giving road users the capability and willingness to use roads and vehicles safely. Several KPIs are introduced:

- KPI for speed: Percentage of vehicles travelling within the speed limit as at higher speed crashes cause far more damage than lower speed ones

- KPI for sober driving: Percentage of drivers driving within the legal limit for blood alcohol content (BAC). The magnitude of the influence of alcohol is difficult to ascertain, thus, the Commission established a maximum permitted blood alcohol content (BAC) of 0.5 % for the general driving population
- KPI for driver distraction: Percentage of drivers not using a handheld mobile device
- KPI for protective equipment: Percentage of PTW riders and of cyclists wearing a protective helmet
- KPI for the use of safety belts and child restraint systems: Percentage of vehicle occupants using the safety belt or child restraint system correctly

4) Ensuring fast and effective emergency response by reducing the transport time between the crash and the arrival of emergency medical services and including qualified personnel for the initial medical treatment provided after a crash. KPI for post-crash care: Time elapsed in minutes and seconds between the emergency call following a crash resulting in injuries and the arrival at the scene of the emergency services.

2.3 Crash databases

A crash is a collision that occurred or originated on roads open to public traffic in which at least one moving vehicle was involved. Road crashes may include collisions between vehicles (intended as vehicles motor vehicles as well as bicycles), between vehicles and pedestrians and between vehicles and animals or fixed obstacles.

Quality of crash data is essential to improving highway safety at all levels of government. Higher organizational actions and management-level activities to improve the standards for the roadway and network design, enhance public health policies and mitigate driver's injury severity depends on data which is used to identify safety issues, determine highway safety messages, strategic communication campaigns and where to adopt selective law enforcement measures, inform decision-makers of needed highway safety legislation, and also evaluate the impact of highway safety countermeasures (NHTSA, 2017). Since this information is valuable in helping to identify ways of improving safety, care should be taken in their interpretation, collection, and objectivity (EU Directive 2019/1936). Furthermore, collecting high-quality data is vital for the full implementation of the 2030 Agenda for Sustainable Development goals and targets through agreed indicators. The issue is far from being completely addressed. As matter of fact, a recent study found that analysts spend the most time cleaning and organizing data instead of building algorithms, exploring data, and doing predictive analysis (CrowdFlower, 2016).

However, crash datasets tend to extremely suffer from the inaccuracy linked with data collection. "Crash underreporting is the rule rather than the exception" has been said by Helvik et al., 2009. Thus, there is an ever-growing demanding change in the method used in reporting routines.

Crash databases are usually built by using police reports and comprising information such as the status of the crash, driver's information, road segment detail, environmental factors, and traffic condition. However, the basic information provided by crash databases suffer from the non-uniformity among countries, among the different states and local jurisdictions in the same country. These issues make data sharing and comparisons difficult and it is to this aim that many researchers have been proposing data improvement. Among them, Montella et al. (2012) proposed to improve the police crash report form by including general information retrieved from medical records that can be used to determine the severity of the crash (the same proposal was also requested by Watson et al., 2015) and the national database with all information collected by the police. In particular, it should contain (a) the crash sketch, showing the main features of the crash site, the movement of vehicles and impact between vehicles and objects; (b) the crash site pictures; (c) the crash narrative with a specific form that includes the manoeuvre of each traffic unit before the crash, the sequence of events of each traffic unit and the environmental and road circumstances; (d) the person violation codes; (e) the injury status of all people involved in the crash (also un-injured persons); (f) the use of safety devices of all the peoples; and (g) the seating positions of all occupants. To check the consistency of the data related to the injury severity, the authors recommend introducing both in the police form as well as in the ISTAT database the new field "number of occupants of each traffic unit".

To encourage greater uniformity and consistency, NHTSA and the Governors Highway Safety Association (GHSA) created the Model Minimum Uniform Crash Criteria (MMUCC) which is a practical guideline with a minimum set of variables to help State and local agencies with the vehicle crash data elements and attributes they should consider collecting information. The first version of MMUC dates back to 1998, then it has been updated four times — in 2003, 2008, 2012, and the last version in 2017 (NHTSA, 2017).

Recently, aiming at eliminating the chances of mistakes caused by illegible handwritten reports and reducing the opportunity for coding errors, some researchers proposed to develop and use a crash database where all components can be integrated automatically (Imprialou & Quddus, 2019) or, in the meanwhile that automated vehicles spread, an electronic form recorded through the use of an advanced GPS-applications which enable police officers to capture and upload crash data from the roadside in real time. For instance, Montella et al. (2019) developed the web application ReGis (Italian acronym of Crash Data Collection, Processing and Analysis) system in Italy. In the UK, the government developed the centralised CRaSH (Collision Recording and Sharing) system to record road traffic crashes and since 2016 the system has been adopted by approximately half of the English police forces. In the US, it was developed the TRACS (Traffic and Criminal Software) state-wide traffic data collection software.

Crash severity definition

Crash severity is defined according to the person involved in the crash who suffered the most serious outcome. Crash severity is also the most accurate indicator of the societal harm (economic as well as social) caused by road crashes. Indeed, when a crash occurs, it results in costs made up of two components: material costs (e.g. damages to vehicles, administrative costs, and medical costs) and immaterial costs (e.g. shorter lifetimes, suffering, pain and sorrow). Even though market prices can be used to calculate material costs, there are no such market prices for immaterial costs (European Commission, 2019b). However, the definition of the different severity levels used by crash investigation professionals is not consistent among all countries and several differences still exist.

No discrepancy exists for fatal crashes where a common definition of fatal injury in road traffic crashes is: “Any person who was killed outright or who died within 30 days as a result of the crash” as given in the Convention of Road Traffic (Vienna, 1968). On the other hand, there is not a commonly accepted and shared definition for serious injuries among countries.

In the United States, road agencies use the KABCO scale to define personal injury severity in crash reports filed by investigating police officers. The scale was ordinal and quantitatively categorizes crashes from the highest to lowest levels of injury severity. It has been defined by the National Safety Council and includes five different levels of severity defined by the National Highway Traffic Safety Association in the 4th edition of the Model Minimum Uniform Crash Criteria (DOT, 2012) and aims to provide a classification of the crash severity for the relative user without having to search through the person level records. The five attributes of crash severity are the following:

- K = Fatal Injury;
- A = Suspected Serious Injury;
- B = Suspected Minor Injury;
- C = Possible Injury;
- O = No-Injury – Property-Damage-Only.

Hence, at each person involved in a crash, a degree of severity is assigned using this scale and the most injured occupant defines the severity of the crash. However, the KABCO severity level is assigned by an investigating police officer and is based on its judgement, rather than a health professional with medical expertise. Thus, investigating officers are subject to the necessary training which can help officers in the interpretation of what they see at the crash scene. It is noted that an injury that turns out to be fatal may be coded as A, B or C up to 30 days after the crash.

The Association for the Advancement of Automotive Medicine introduced the Abbreviated Injury Scale (AIS) to identify the injury severity level of the person involved in the crash based on anatomic disruption upon an in-hospital clinical assessment (AAAM, 2005). According to the AIS trauma scale, using existing medical records, each individual anatomical injury is identified with a code with a severity score ranging from 1 (minor) to a maximum value equal to 6 (life-threatening). In so doing, each injured person has an AIS vector of codes addressing all major body regions. The possible codes that can be used for each injury are the following:

- 1 = Minor;
- 2 = Moderate;
- 3 = Serious;
- 4 = Severe;
- 5 = Critical;

- 6 = Maximal.

To easily compare the injury severity across individuals or from crash to crash, recently, the EU has introduced the Maximum Abbreviated Injury Scale (MAIS). The MAIS, which is the maximum AIS across all body regions, aims to provide a common definition of crash severity based on a combination of the vector of severity scores into a single metric for any patient. Different meaningful levels for MAIS are MAIS2+ (MAIS≥2, at least moderate injury), MAIS3+ (MAIS≥3, at least serious injury), MAIS4+ (MAIS≥4, at least very serious injury), MAIS5+ (MAIS≥5, at least critical injury). However, the most utilized is MAIS 3+ which has been fixed as the cut-off value for defining serious injury. Thus, a casualty that sustains an injury with an AIS score of 3 or higher is classified as clinically seriously injured. Because of the difficulties associated with a common definition for non-fatal injuries, since January 2013, the High Level Group on Road Safety, representing all EU Member States, established the definition of serious injuries as road casualties with an injury level of MAIS ≥ 3 (Weijermars et al., 2018). The High Level Group also identified three main ways the Member States can collect data on serious injuries depending on the available data:

1. by applying a correction factor on police data;
2. by using hospital data alone;
3. by using linked police and hospital data.

Currently, not every EU country has introduced the MAIS3+ to classify serious injuries due to the main problems related to the limited access to hospital discharge data due to privacy regulations.

One criticism of the MAIS trauma scale is that the MAIS3+ score does not provide differences between patients with several serious injuries to several body regions and those with more localised injuries.

As an alternative, Baker et al. (1974) developed the Injury Severity Score (ISS). The ISS is the sum of the squares of the highest AIS scores in three different most severely injured body regions (out of 6 regions). ISS ranges from 0 to 75. ISS=75 indicates a fatality or no chance of survival and is assigned if even just one body region has been scored with an AIS of 6. An ISS16+ is the cut-off value used to define seriously injured occupants.

Among the various definition of a “serious” crash, general wisdom has it that either MAIS3+ or ISS16+ gives a more precise estimate of the serious injury severity of a crash than K and A crashes on the KABCO scale (Flannagan et al., 2013, Ivan and Konduri, 2018) as KA tends to overestimate the number of serious injuries by about 3 times. However, in some countries, such as Sweden, the use of MAIS is still too far from being eligible to be used in crash severity analyses. The common definition of a serious injury is related to the ISS trauma scale.

Table 13 and Table 14 summarized the current European countries' progress in collecting data on seriously injured based on MAIS3+ definition. In-depth information was provided by ETSC in the 15th Annual Road Safety Performance Index (PIN) Report (ETSC, 2021).

Table 13 – European countries' progress in collecting data on seriously injured based on MAIS3+ (Source: ETSC, 2021), part A.

AT	In 2015, using hospital data, the number of MAIS3+ injuries was estimated for the first time for the year 2014 and has been continued for all years thereafter. Time series available starting 2010.
BE	Belgium is fine-tuning the procedure of MAIS3+ estimation using correction factors applied to police data and using of hospital data.
BG	The only source is Police records.
CY	For 2017 and 2018, the data were provided based on MAIS3+. For 2019 and 2020, it is unpredictable when the number will be calculated, because of the COVID19 crisis.
CZ	The implementation of MAIS3+ will be applied potentially in 2022. Negotiations between the Ministry of Interior and the Ministry of Health is under way.
DE	MAIS3+ injured persons estimation is calculated based on GIDAS data, data from the German Trauma Register, and data from the official crash statistics.
DK	No systematic linkage between police and hospital data. Denmark is working on a process to convert ICD diagnose codes into AIS and MAIS.
EE	In 2019 Estonia tried to test EU proposed ICD - AIS conversion tool with doubtful results. Further work depends on the initial data quality and convention tool (AAAM) updates.
ES	Since 2010.
FI	Since 2014.
FR	MAIS3+ injury estimation are currently being evaluated.
EL	Hospitals do not systematically collect data on the injury severity of road casualties.
HR	Link between police and hospital is based on the law. However, MAIS3+ is no available.
HU	Link between police and hospital data is not provided. However, the National Healthcare Services Centre started to upgrade the information system.
IE	MAIS3+ estimation by conversion tables made available by the EC provides doubtful results. Collecting serious injuries using a medical definition will be a priority within the Road Safety Strategy 2021 – 2030.
IT	Link between police and hospital data is not provided. MAIS3+ has been adopted for coding the level of injury based on hospital data. An estimate of the number of seriously injured has been calculated since year 2012 according to the conversion tables made available by EC.

AT = Austria, BE = Belgium, BG = Bulgaria, HR = Croatia, CY = Cyprus, CZ = Czechia, DK = Denmark, EE = Estonia, FI = Finland,FR = France, DE = Germany, EL = Greece, HU = Hungary, IE = Ireland, IT = Italy.

Table 14 – European countries' progress in collecting data on seriously injured based on MAIS3+ (Source: ETSC, 2021), part B.

LU	MAIS3+ will be used in the near future.
LV	Latvia is planning to start registered serious injuries based on MAIS3+ from January, 2022.
LT	Since 2014.
MT	MAIS3+ conversion process is still ongoing.
NL	Data on MAIS3+ already available for the period 1993-2018.
PL	Poland converted data from 2013 and 2014 according to the EU recommendations. However, recently the work on MAIS 3+ has been stopped due to potential errors and doubtful results. Unfortunately, due to a lack of financing, Poland could not launch a national project to develop a methodology for assessing the severity of injuries of road accident victims according to the MAIS 3+ scale.
PT	Since 2015. A new procedure to collect the police data while preserving the victim's privacy has been established.
RO	Since 2021.
SE	Since 2007.
SI	Slovenia tried experimental linking between police and hospital data. However, MAIS3+ data are incomplete and not ready for publication and still under discussion.
SK	Under discussion.
CH	Linking of health and police data has started in 2014. This allows to code the recommended maximum AIS score based on ICD-10.
IL	Since 2013 police data is linked with hospital data. Any casualty found in both sources, their injury severity is defined by MAIS. If the casualty was not found in the hospital data, their injury severity is defined by the police. Seriously injured is defined by MAIS 3+ or hospitalized for a period of 24 hours or more, not for observation only.
NO	Under consideration.
RS	Road Traffic Safety Agency intends to introduce MAIS3+ definition of serious injuries in road traffic crashes in the next period.

LV = Latvia, LT = Lithuania, LU = Luxembourg, MT = Malta, NL = The Netherlands, PL = Poland, PT = Portugal, RO = Romania, SK = Slovakia, SI = Slovenia, ES = Spain, SE = Sweden, GB = Great Britain, IL = Israel, NO = Norway, RS = Serbia, CH = Switzerland.

2.4 Statistical methods

To date, to better face the public concern related to pedestrian crash severity, different analytical methods and methodological approaches have been used over time attempting to identify roadway features, driver and pedestrian behaviours, and other contextual contributory factors associated with pedestrian crashes to gain such an understanding. To this end, analysis and review of existing literature about historical and recent methodologies used for crash severity analysis are provided in Table 15 - Table 17.

Analytical methods for crash severity prediction may be grouped into two main categories, namely econometric models and machine learning (ML) algorithms.

The econometric models, also reckoned as discrete choice models, are widely used in crash severity analyses. These models use the theoretical utility U_{ij} which, in the context of road safety applications, represents the propensity for a crash i of being recorded with severity level j, following the expression reported below:

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \sum \beta_j x_{ij} + \varepsilon_{ij} \tag{1}$$

where V_{ij} is the systematic component able to capture the contribution of independent variables to a crash event and its crash severity (with x_{ij} being the column vector of the observable characteristics or independent variables that affect the crash outcome and β_j the column vector of the estimable parameters) whereas ε_{ij} is a disturbance term. Many studies reported the value of β_j estimated as well as the $\exp(\beta_j)$. The factor $\exp(\beta)$ is the odds ratio (OR) and indicates the relative amount by which the odds of the outcome increases (OR >1) or decreases (OR <1) when the value of the corresponding indicator variable is 1.

If crash severity is a three-level variable (i.e., as for Great Britain and Sweden), it is well adaptable to econometric models with their multinomial formulation as well as the ordinal one. Therefore, each level of crash severity is linked with: 1) an increasing severity of the most seriously injured person involved in the crash, and 2) an increasing cost in terms of human, medical and damage costs, involving loss of life years and quality of life. Thus, crash severity has an ordinal nature which could be addressed by performing the analysis with econometric model ordered formulation. In this research, both unordered and ordinal logit models were developed. Furthermore, both unordered and ordered models were used in the standard formulation with fixed parameters as well as in the formulation with random parameters (Figure 4). Random-parameter models allow the effect of independent variables to vary across different observations (crashes in this research).

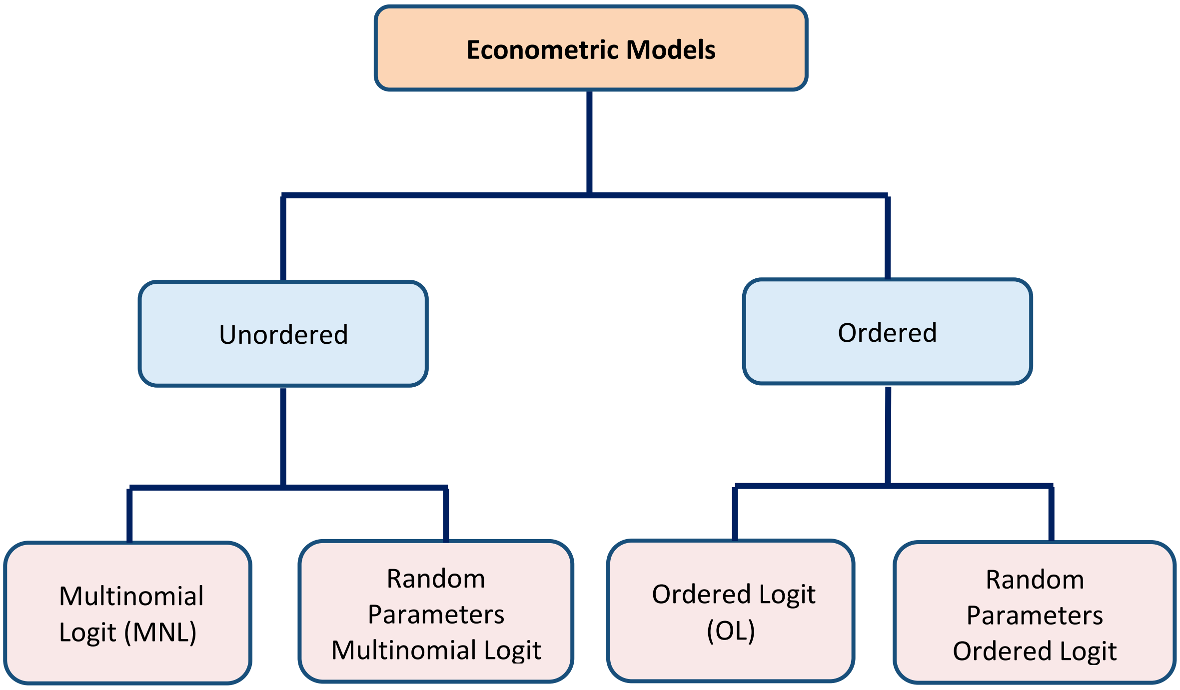


Figure 4 – Most popular Econometric models.

The most used econometric model to investigate and identify significant factors contributing to pedestrian-vehicle crash severity are the binary logit or the multinomial logit (MNL) (e.g., Casado-Sanz et al., 2019; Chen and Fan, 2019; Hanson et al., 2013; Rothman et al., 2012, as reported in Table 15) which used the unordered response structure. Although multinomial logit has undoubtedly provided important insights, methodological limitations that are not fully understood could have affected study results leading to erroneous inferences and biased crash predictions (Mannering et al., 2016; Savolainen et al., 2011). Indeed, over the past decade, in the context of road safety analyses, many studies have highlighted the importance of accounting for the unobserved heterogeneity which is the presence of countless factors which are unlikely to be observed by the data analyst but that can influence crash occurrences and the resulting injury severities. Thus, several methodological approaches have been performed to gain more precise estimations by explicitly accounting for observation-specific variations in the effects of explanatory variables (Ahmadi et al., 2018; Malyshkina and Mannering, 2010; Mannering et al., 2016, 2020; Milton, 2006; Washington et al., 2011). Among them, the random parameters (or simply mixed) model implies the strong assumption of considering the unobserved individual-specific heterogeneity to be completely unrelated to the explanatory-variable vector. Indeed, for the mixed models, the parameter effects can vary across individual crashes ranging from negative to positive and of varying magnitudes (Milton, 2006).

Recognizing the ordinal nature of crash severity data, other studies have been conducted by performing ordered-response models and setting specific thresholds for the probability of injury severity to various alternatives (Yasmin et al., 2014; Pour-Rouholamin & Zhou, 2016). Thus, among the most popular discrete choice approaches, discrete ordered-probability methods (ordered logit and probit models), have shown great appeal. Yamamoto et al. (2008) further suggested that traditional unordered models may provide an unbiased estimate of the parameters, especially in the case of missing data such as under-reporting. Even though the ordered-response model may have the advantage that its data generation process is more consistent with the ordinal nature of the injury severity variable, many researchers (e.g., Paleti et al., 2010) pointed out that a limitation of the traditional ordered-response structure is that it imposes a certain kind of monotonic effect of independent variables on injury severity levels. Interestingly, the authors suggested a new avenue for research: consider generalizations of the traditional ordered-response models and examine their predictive ability and behavioural validity. Indeed, the strong restrictions on the fixed threshold across observations represent a critical component in the application of the standard ordered response models (Abay, 2013; Eluru et al., 2008; Yasmine et al., 2014). Therefore, the mixed ordered response logit model for analysing crash data can generalize the standard ordered response models by allowing the flexibility of the effects of covariates on the threshold value for each ordinal category and capturing the heterogeneous effects, both of which cannot be by traditional probability models (Mokhtarimousav et al., 2020; Ye and Lord, 2011; Srinivasan, 2002). However, very few studies implemented the random parameter ordered logit (Yasmin et al., 2014; Pour-Rouholamin & Zhou, 2016; Mokhtarimousavi et al., 2020) and probit models (Mokhtarimousavi et al., 2020).

Both ordered and unordered models have their benefits and limitations, and the choice of one method over the other is governed by the availability and characteristics of the data and involves considering trade-offs (Cerwick et al., 2014).

Table 15 – Summary of prior research on pedestrian crash severity – Fixed econometric models.

Authors	Model	Study period	Total crash data
Moudon et al., 2011	Binary ordered/ unordered Logit	1999-2004	711 on route states
Moudon et al., 2011		2000-2004	2,351 on city streets
Hanson et al., 2013	Binary Logit	2007-2009	6,353
Zhang et al., 2014		2006-2010	6,967
Olszewski et al., 2015		2007-2012	18,850
Rothman et al., 2015		2000-2009	23,428
Noh et al., 2018		2008-2015	79’078
Ghasedi et al., 2021		2017-2019	1,061
Damsere-Derry et al., 2010	Multinomial Logit	2002-2006	812
Rothman et al., 2012		2000-2009	9,575
Casado-Sanz et al., 2019		2006-2016	1’535
Chen & Fan, 2019		2005-2012	3’553
Yasmin et al., 2014	Ordered Logit	2002-2006	7,354
Pour-Rouholamin & Zhou, 2016		2010-2013	19,361
Lee & Abdel-Aty, 2005	Ordered Probit	1999–2002	7’000
Clifton et al., 2009		2000-2004	4’695

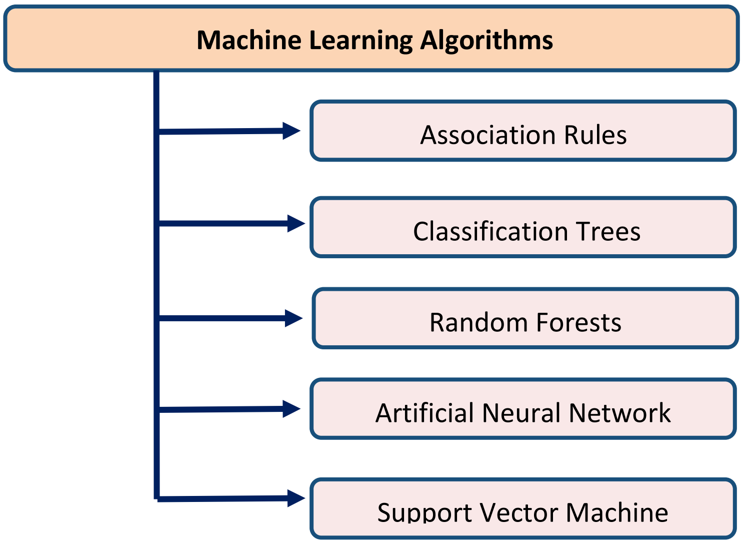
Table 16 – Summary of prior research on pedestrian crash severity – Mixed econometric models.

Authors	Model	Study period	Total crash data
Zhai et al., 2019	Random Parameter Binary Logit	2015	2’794
Tulu et al., 2017	Random Parameter Multinomial Logit	2009-2012	6,208
Kim et al., 2010		1997-2000	5’808
Aziz et al., 2013		2002-2006	7’354
Islam & Jones, 2014		2006-2010	1,463
Haleem et al., 2015		2008-2010	7,630
Eluru et al, 2008	Random Parameter Ordered Logit	2004	2’944
Yasmin et al., 2014		2002-2006	7,354
Pour-Rouholamin & Zhou, 2016		2010-2013	19,361
Abay et al., 2013	Ordered Logit Random Parameter Ordered Logit Multinomial Logit Random Parameter Multinomial Logit	1998-2009	4,952
Mokhtarimousavi et al., 2020	Random Parameter Ordered models: Logit and Probit	2010-2014	10’146

The econometric models suffer fundamental limitations, such as the presumption of crash data distribution and their restrictions on the linear relationship between severity outcomes and explanatory variables. Furthermore, it is also well-known that no injury or minor injury crashes are very rarely reported to police (Imprialou et al., 2019; Washington et al., 2020; Yamamoto et al., 2008; Ye and Lord, 2011) and an outcome-based model may result in biased parameter estimates when traditional statistical estimation techniques are used limiting the ability to manage road safety. Another downside of the traditional statistical models is related to their difficulties in handling and processing very large amounts of data, so that, in the last few years data-driven methods have been applied to crash analysis attempting to overcome the issue.

Among the many algorithms available in the literature, a group of five popular supervised machine learning algorithms, namely Classification Tree (CT), Random Forest (RF), and Support Vector machine (SVM) have been recently used to predict injury severity (Figure 5).

Figure 5 – Most popular Machine Learning algorithms.



Free from a priori parametric assumptions typical of econometric models (Mannering, 2020), data-driven methods, also known as ML algorithms, include association rules (AR), classification trees (CT), random forests (RF), artificial neural networks (ANN), and support vector machine (SVM) which are the five machine learning algorithms used in this research. AR discovery (also known as the supervised association mining technique) has been widely used to discover patterns from the crash database (Montella, 2011) and, recently, has been applied to vehicle-pedestrian crashes (Besharati et al., 2017; Das et al., 2018; Sivasankaran et al., 2020). CTs have been already developed to uncover patterns influencing crash severity in countless papers. Among them, Montella et al. (2012, 2020) used classification trees to predict crash severity for different road users finding the tree structure effective in providing a clear understanding of the phenomenon under study and in some papers researchers used CTs to investigate factor contributing to pedestrian crash severity (Montella et al., 2011; Li et al., 2017; Ospina-Mateus et al., 2019). Recently, other researchers implemented the RF tool which exhibits more stable outputs than the CT since it considers an ensemble of trees instead of one (Li et al., 2017; Komol et al., 2021). Another tree-structure algorithm is the ANN tool. Recently, the tool has been implemented by Ospina-Mateus et al. (2019), Mokhtarimousavi et al. (2020) and Ghasedi et al. (2021) for vehicle-pedestrian crashes. Among data-driven methods there is an increasing interest also in using the SVM tool to investigate patterns contributing to pedestrian crash severity (Ospina-Mateus et al., 2019; Komol et al., 2021) due to the straightforward algorithm ability in providing the goodness of fit and better prediction performance than the other traditional methods. Despite this advantage, the tool is usually defined as a black box in nature and the output interpretation may result difficult as well as the consequent safety countermeasures to implement.

Table 17 – Summary of prior research on pedestrian crash severity – Machine Learning algorithms.

Authors	Model	Study period	Crash data
Montella et al., 2011	AR CART	2006-2008	56'014
Das et al., 2019	AR	2004-2011	11'503
Besharati et al., 2017	AR	2009-2012	34,178
Sivasankaran et al., 2020	AR	2015-2016	3,416
Li et al., 2017	CART RF	2013	14'174
Mokhtarimousavi et al., 2020	ANN	2010-2014	10'146
Ghasedi et al., 2021	ANN	2017-2019	1,061
Komol et al., 2021	KNN SVM RF	2013-2019	21'158
Ospina-Mateus et al., 2019	CT SVM ANN Naïve Bayes	2016-2017	10,053
Meocci et al., 2021	Gradient Boosting	2014-2018	101,030

To sum up, despite the great efforts that have been demonstrated by researchers in the attempt to analyse crash severity, the debate about the most appropriate method is still on and just a handful comparisons among the methods have been performed (Ahmadi et al., 2018; Mokhtarimousavi et al., 2020; Zhang et al., 2018). However, such an in-depth comparative analysis among econometric models and ML tools applied on the same database (and on more databases) is still lacking in the literature. Therefore, in this research I estimated multinomial logit (MNL), ordered logit (OL), random parameters multinomial logit (RPMNL), and random parameters ordered logit (RPOL) as econometric models and AR, DT, RF, ANN, and SVM as ML algorithms. Then, the results of the methods were compared by their classification performances through the evaluation of performance metrics. A further qualitative comparison was provided in terms of potential contributory factors identified by each method with the aim of detecting similarities and dissimilarities among the models as well as finding meaningful insights about pedestrian crash severity patterns and their interdependencies.

2.5 Results of prior research

Road crashes are complex, rare, and random events. Complex as they involve a variety of human responses to external stimuli caused by failed interactions between vehicles, roadway features, and environmental conditions. Rare implied that crashes represent only a very small share of the total number of events that occur on the transportation system due to the interaction among all the road users and the road environment. Random means, instead, that crashes occur as a function of a set of events influenced by several factors, which are partly deterministic (they can be controlled) and partly stochastic (random and unpredictable). Thus, when a crash occurred, some patterns could be due to the chance alone or, in many other situations, could be strongly connected to each other.

The safety effects of factors contributing to road crashes can be sometimes easily conceived and relatively consistent (i.e., the increase in the probability of observing a road crash in presence of considerable exposure and high travel speeds). However, there are other contributory factors which impact on crash severity is not easy to capture. Furthermore, the impacts of these factors can vary on different crash severity levels according to the road users involved in the crash (drivers, pedestrians, cyclists, …) or the area of crash (i.e., urban or rural roads). Bearing this in mind, the present literature review will focus on which factors contributing to the most serious pedestrian crashes have been introduced and investigated in previous studies. The large number of fatal and severe injury vehicle-pedestrian crashes provides evidence that the safety of pedestrians is a key issue when improving traffic safety. To date, indeed, a significant amount of pedestrian safety research has been undertaken. A summary of prior research on pedestrian injury severity analysis is provided from Table 18 to Table 20 aiming to sum up information on the crash severity (dependent variable), and all characteristics/factors considered in the analysis (including road characteristics, roadway design and land use attributes, vehicle characteristics, environmental factors, crash characteristics, driver characteristics, and pedestrian characteristics). A resume of the method framework employed in each study was provided in the next paragraph (see paragraph 2.4).

Table 18 – Main characteristics investigated in prior research (Part A).

Authors	Crash severity levels	Characteristics considered in the analysis					
		Roadway	**Vehicle**	**Environment**	**Crash**	**Driver**	**Pedestrian**
Abay et al., 2013	1. Killed or fatal 2. Severe injury 3. Slight injury 4. No injury/no casualty	X	X	X	X	X	X
Aziz et al., 2013	5. Fatal 6. Severe injury 7. PDO and possible injury	X	X	X	X		X
Besharati et al., 2017	1. Fatal 2. Injury	X		X		X	X
Casado-Sanz et al., 2019	1. Fatal 2. Severe injury 3. Slight injury	X		X	X	X	X
Chen & Fan, 2019	1. Fatal 2. Injury class 1 3. Injury class 2 4. Injury class 3 5. No injury	X	X	X		X	X
Clifton et al., 2009	1. Fatality 2. Injury 3. No injury	X	X	X			X
Das et al., 2018	1. Fatal 2. Injury 3. PDO	X		X	X	X	X
Demsere-Dery et al., 2010	1. Fatal 2. Hospitalized injury 3. Slight injury	X	X	X		X	X

Table 19 – Main characteristics investigated in prior research (Part B).

Authors	Crash severity levels	Characteristics considered in the analysis					
		Roadway	Vehicle	Environment	Crash	Driver	Pedestrian
Eluru et al., 2008	1. Fatal injury 2. Incapacitating injury 3. Non-incapacitating injury 1. No injury	x	x	x	x	x	x
Kim et al., 2010	2. Fatal injury 3. Incapacitating injury 4. Non-incapacitating injury 5. Possible/no injury	x	x	x	x	x	x
Hanson et al., 2013	1. Fatal 2. Other	x		x			x
Islam & Jones, 2014	1. Major injury 2. Minor injury 3. Possible/no injury	x		x	x		x
Lee & Abdel-Aty, 2005	1. Fatal 2. Incapacitating injury 3. Non-incapacitating injury 4. Possible injury 5. No injury		x	x			x
Li et al., 2017	1. Fatal 2. Severe injury 3. Slight injury			x		x	x
Mokhtarimousavi et al., 2020	1. Severe injury 2. Minor injury 3. No injury		x		x	x	x

Table 20 – Main characteristics investigated in prior research (Part C).

Authors	Crash severity levels	Characteristics considered in the analysis					
		Roadway	Vehicle	Environment	Crash	Driver	Pedestrian
Olszewski et al., 2015	1. Fatal 2. Other	X		X			X
Ospina-Mateus et al., 2019	1. High level of injury 2. Low level of injury			X	X	X	X
Pour-Rouholamin et al., 2016	1. Severe injury 2. Minor injury 3. Possible/no injury	X	X	X		X	X
Rothman et al., 2015	1. Fatal/severe injury 2. Minor/minimal/no injury	X					X
Sivasankaran et al., 2020	1. Fatal/grievous 2. No/simple injury	X		X	X	X	X
Tulu et al., 2017	1. Fatal 2. Serious injury 3. Slight injury	X	X			X	X
Zhai et al., 2019	1. Fatal/severe injury 2. Slight injury	X		X			X
Montella et al., 2011	1. Fatal 3. Injury	X	X	X		X	X
Moudon et al., 2011	1. Fatal/high injury 4. Low/no injury	X	X	X		X	X
Noh et al., 2018	1. Severe injury 5. Non severe injury	X	X	X		X	X

The analysis of previous research revealed that in earlier studies on pedestrian crash severity analysis, the dependent variable crash severity mostly ranged between two (with a binary response, generally fatal/injury) or five (fatal, incapacitating injury, non-incapacitating injury, possible injury, no injury) levels of severity. In many cases, it was classified on three levels (generally fatal/serious injury/slight injury). In very few cases, authors used a more than three level response variable. Furthermore, in many cases, fatal and serious injury crashes were merged together in a single level due to the extremely presence of imbalanced data (fatal and serious injury covered a small share of the total crash consequences) which may create bias in statistical models. However, the choice related to how to consider the crash severity in the analysis is often also a mere consequence of how the variable is collected in police records (i.e., Italian database provides the variable crash severity as a binary variable: fatal or injury without any information on the level of severity sustained by injuries).

The investigation of factors associated with pedestrian crashes should consider the existence of pedestrian, driver, vehicle, and environmental factors which may have caused an increase in pedestrian fatalities and severe injuries in recent years. When the interactions between these factors and severity are co-considered and co-investigated, the injury causes as well as the related solutions are better identified (Theofilatos & Yannis, 2015). Among the reported studies, only three previous research (Abay et al., 2013; Eluru et al., 2008; Kim et al., 2010) examined variables from all factor categories in their empirical analysis. The tables below provide a resume of the main results achieved. Red up-arrows stand for indicators which presence lead to an increase in pedestrian crash probability crash severity, while the indicators associated with green down-arrows result in a reduction in this probability. If a factors has both red and green arrows, it does mean that that pattern was estimated as random by the model implemented in the relative study (i.e., random parameter multinomial logit, random parameter binary logit, random parameter ordered logit, …).

Roadway characteristics have been widely investigated in pedestrian crash severity analysis, Table 21. When crashes happen, usually drivers are blamed for the mishap. However, error is part of the human condition. Hence, if drivers consistently and repetitively fail at certain locations, it then means that the problem lies not with them, but with the road itself. Even though road users try to drive or behave in a safe way, driving is a complex task and the environment is not designed to prevent such errors occurring or in forgiving potential driver errors. Moreover, because crashes are not evenly distributed throughout the road network, locations with a considerable number of crashes are a clear indication that there are other factors involved, besides driver error, which are characterized by the road itself (Lamm et al., 2007).

Table 21 – Main results for roadway characteristics in pedestrian crash severity analysis.

Roadway characteristic	Authors	Fatal crashes	Serious injury
Rural area	Lee and Abdel-Aty, 2005	↑	↑
	Montella et al., 2011 Olszewski et al., 2015 Besharati et al., 2017	↑	
Urban area	Pour-Rouholamin & Zhou, 2016 Chen and Fan, 2019	↓	↓
Straight-but-not-level roads	Aziz et al., 2013	↑	↑
Straight- level roads	Das et al., 2018	↑	↑
Straight road section	Demsere-Derry et al., 2010	↑	
Number of lanes	Meocci et al., 2021	↑	
Single lane	Aziz et al., 2013	↓	
Two-way divided	Kim et al., 2010 Olszewski et al., 2015	↑	
Lane width < 3.25 m Lane width between 3.25-3.75 m	Casado-Sanz et al., 2019	↑	↑
Shoulder width < 2.5 m,	Casado-Sanz et al., 2019	↓	↓
Speed limits above 50 mph	Chen and Fan, 2019	↑	↑
Speed limits above 50–60 mph	Eluru et al., 2008	↑	
Speed limits above 40–70 mph	Li et a., 2017	↑	↑
Curve	Kim et al., 2010		↑
	Montella et al., 2011	↑	
	Chen and Fan, 2019	↑	↑
Not at junction	Demsere-Derry et al., 2010	↑	↑
Intersections without signal control device	Aziz et al., 2013	↓	
	Lee and Abdel-Aty, 2005	↑	↑
Signalized intersection	Eluru et al., 2008	↓	
Land use: vacant land	Besharati et al., 2017	↑	
Presence of pedestrian attractors in the study area	Moudon et al., 2011	↑	

Road characteristics include the area, the road design and all conditions related to the alignment and to the road traffic control devices. Among them, many prior studies investigated the intersection design. Indeed, in many cities, streets may have sidewalks and other protective pedestrian devices to walk along them whereas no protection or sufficient devices are provided for pedestrian crossing the street. Furthermore, some intersections may be more complicated than others and this is what previous studies analysed. Other road factors relate to the neighbourhood land use, the presence of schools, colleges, or other activities with a significant attractive power for pedestrians.

The vehicle type (Table 22) is crucial in crash severity analysis and it further play a key role in severity of pedestrian crashes. In such circumstances, pedestrians do not have any kind of protection able to absorb the impact force.

Table 22 – Main results for vehicle type in pedestrian crash severity analysis.

Vehicle Type	**Authors**	**Fatal crashes**	**Serious injury**
Truck/Bus	Kim et al., 2010 Aziz et al., 2013 Noh et al., 2018	↑	
Truck/Bus/Van	Lee & Abdel-Aty, 2005 Clifton et al., 2009 Pour-Rouholamin & Zhou, 2016	↑	↑
Truck	Montella et al., 2011 Besharati et al., 2017 Chen and fan, 2019 Zhai et al., 2019	↑	
	Tulu et al., 2017	↓/↑	↑
SUV	Eluru et al., 2008	↑	
Car	Moudon et a. 2011	↑	↑
PTW	Noh et al., 2018	↓	
	Chen & Fan, 2019	↑	↑

Environmental factors regard the pavement conditions, the lighting, the weather, and time of the day at which the crash occurred. The information is related to conditions of the road and environment which were unique to the time and location of the crash. Some factors, such as the weather, cannot be controlled. However, there are other factors, road pavement conditions and lighting conditions just as examples, which influence crash severity and, if identified their impact, it could be minimized by best practices in terms of road safety policies by government. As far as the lighting is concerned, many studies (as reported in Table 23) analysed different conditions: dark roads with the complete absence of lighting, or roads which are dark even in presence of road illumination. Other researchers focused on night-time, in opposition with day-time and natural lighting. Few studies introduced the variable season to understand if pedestrian crashes severity may be somehow related to seasonal effects.

Table 23 – Main results for environmental factors in pedestrian crash severity analysis.

Environmental condition	Authors	Fatal crashes	Serious injury
Wet pavement	Aziz et al., 2013	↓	↑
	Chen & fan, 2019	↓	
Dark roads	Aziz et al., 2013	↓/↑	
	Islam & Jones, 2014 Besharati et al., 2017 Li et al., 2017	↑	
Dark roads (lighted and unlighted)	Lee & Abdel-Aty, 2005 Kim et al., 2010 Pour-Rouholamin & Zhou, 2016 Das et al., 2018 Chen & Fan, 2019	↑	↑
Night-time	Eluru et al., 2008 Damsere-Derry et al., 2010 Montella et al., 2011 Olszewski et al., 2015 Noh et al., 2018	↑	
	Li et al., 2017	↑	↑
Dusk and dawn	Chen & Fan, 2019	↑	
Clear weather	Islam & Jones, 2014	↓/↑	
	Sivasankaran et al. 2020	↑	↑
Heavy rain	Lee and Abdel-Aty, 2005	↑	↑
	Zhai et al., 2019 Ghasedi et al., 2021	↑	
Inclement weather	Kim et al., 2010	↓	
Summer season	Olszewski et al., 2015	↑	
	Pour-Rouholamin & Zhou, 2016	↑	↑
Weekend	Tulu et al., 2017	↑	↑

Crash characteristics include an interesting set of variable which have been unfortunately investigated only in very few prior research. This is also due to the lack of reporting of this information in crash records.

Table 24 resume prior efforts in this direction. Mainly, prior results provided that a frontal impact between a pedestrian and a vehicle have the highest probability of resulting serious or fatal. In other studies, the point where the pedestrian cross the road has been considered revealing that especially unsignalized intersections may have a pivotal role in the severity of pedestrian crashes. Some research (i.e., Aziz et al., 2013) found a random effect in crash severity for pedestrian crossing at signalized intersections considering that the severity of crashes at intersection may also depend on the behaviour of certain drivers and their driving attitude.

Table 24 – Main results for crash characteristics in pedestrian crash severity analysis.

Crash characteristics	Authors	Fatal crashes	Serious injury
Number of pedestrian involved >1	Moudon et al., 2011	↑	↑
Frontal impacts	Eluru et al., 2008	↑	↑
Pedestrian crossing the roadway	Damsere-Derry et al., 2010	↑	↑
Pedestrian crossing at unsignalized intersections	Moudon et al., 2011	↑	↑
Pedestrian crossing at signalized intersections	Aziz et al., 2013	↓/↑	↑
Pedestrian not using crossings	Casado-Sanz et al., 2019	↑	↑
Pedestrian illegally crossing near the pedestrian overpass	Noh et al., 2018	↑	
Walking against traffic	Islam & Jones, 2014	↓	

Road users' factors include users' characteristics and users' behavioural aspects (Table 25). Human behaviours (impaired driving due to alcohol or drug use, speeding, aggressive manoeuvres such as tailgating and inappropriate overtaking) have been introduced in several analysis. However the information is not available in all crash databases. As the crash information are collected by the police, some factors mentioned in the reports may be largely subjective, reflecting the opinion of the reporting police officer and which are not necessarily the result of extensive investigation.

Other factors, instead, are less likely to be recorded since evidence may not be available after the crash event.

In some studies, driving/riding skills, experience, and users' psychophysical state have been also considered revealing, for instance, that unexperienced drivers in possess of driving licence for less than 5 years, are more likely to be involved in the most serious pedestrian-vehicle crashes. Among road factors also age and gender are considered.

Table 25 – Main results for driver characteristics in pedestrian crash severity analysis.

Drivers' factors	Authors	Fatal crashes	Serious injury
Left turning manoeuvre	Aziz et al., 2013		↑
Turning/merging manoeuvre	Kim et al., 2010	↓	↓
Reversing manoeuvre	Kim et al., 2010	↓	↓
Unexpected manoeuvres (Changing lane, overtaking)	Li et al., 2017	↑	↑
Male driver	Kim et al., 2010 Das et al., 2018	↑	↑
Female driver	Besharati et al., 2017	↑	
Diver age under 24	Pour-Rouholamin & Zhou, 2016	↑	↑
Driver age over 25	Chen & fan, 2019	↓	↓
Driver age over 65	Pour-Rouholamin & Zhou, 2016	↓	↓
Driver has been drinking	Kim et al., 2010 Moudon et al., 2011 Pour-Rouholamin & Zhou, 2016	↑	↑
	Eluru et al., 2008	↑	↓
Speeding	Kim et al., 2010 Casado-Sanz et al., 2019 Sivasankaran et al. 2020	↑	↑
	Damsere-Derry et al., 2010	↑	
Negligent driving	Demsere-Derry et al., 2010 Zhai et al., 2019	↑	
Driving experience (under 5 years)	Tulu et al., 2017	↑	

Many studies have also investigated the characteristics of pedestrians in vehicle crashes. Among them, the age and the gender of pedestrians, the influence of alcohol, the behaviour of the pedestrians, and the colour of the clothes they wore when the crash occurred (Table 26).

Table 26 – Main results for pedestrian characteristics in pedestrian crash severity analysis.

Pedestrians' factors	Authors	Fatal crashes	Serious injury
Pedestrian age under 6	Besharati et al., 2017	↑	
Pedestrian age under 15	Pour-Rouholamin & Zhou, 2016	↑	↓
Pedestrian age under 18	Rothman et al., 2015	↑	↑
Pedestrian age 18-30	Casado-Sanz et al., 2019	↓	↓
Pedestrian age over 45	Lee & Abdel-Aty, 2005 Moudon et al., 2011	↑	↑
	Montella et al., 2011 Zhai et al., 2019	↑	
Pedestrian age over 65	Clifton et al., 2009 Rothman et al., 2015 Pour-Rouholamin & Zhou, 2016 Li et al., 2017 Casado-Sanz et al., 2019	↑	↑
Pedestrian age over 65	Tulu et al., 2017	↓/↑	↑
Male pedestrian	Das et al., 2018	↑	↑
	Montella et al, 2011	↑	
Female pedestrian	Lee & Abdel-Aty, 2005 Besharati et al., 2017 Noh et al., 2018	↑	
Alcohol/drug use by pedestrian	Lee & Abdel-Aty, 2005	↑	
Pedestrian wearing reflective clothes	Islam & Jones, 2014	↓	
Pedestrian wearing dark clothes	Besharati et al., 2017	↑	

CHAPTER III ~ METHODOLOGICAL APPROACH

Crashes are strongly influenced by randomness. This source of variation is particularly prominent in small crashes databases. Instead, larger databases tend to reduce this phenomena because of the law of large numbers prevails.

Detecting on the interdependence as well as the dissimilarities among crash characteristics is useful in providing insights on crash causes and suggesting possible improvements on road safety. Overall, many authors focused their studies on crash severity analyses testing different methods and their reliability in order to evaluate the most appropriate engineering countermeasures and policies aimed at reducing the deaths on road.

Last few years have been characterised by a real alarm towards vulnerable users' health and in particular pedestrians. Below, a description of the most used (or the most recent) models pointed out in the literature review is here provided.

3.1 Econometric models

3.1.1 General issues

All the econometric models, which are going to be better described as follow, were estimated by the maximum likelihood stepwise method which implicitly tests the correlation among independent variables. Indeed, to choose a model, the forward stepwise approach begins with a null model and adds terms sequentially until further additions do not improve the fit. From a potentially large set of variables it chooses a subset to include in the model. The most ambitious form of the variable selection relies on the "best-subset" selection, so that the stepwise is a procedure which picks the best model among all 2^G subsets of the G possible groups (Loftus & Taylor, 2014). Once a variable is included in the model, it will remain throughout the process. The variables sequentially added in the model should satisfy some optimality criterion. A common criterion used in stepwise procedure for regression models is to add variables at each step whose partial F-statistic yields the smallest p-value. Variables are included in the model and entered as long as the partial F-statistic p-value remains below a specified maximum, say PIN. The stepwise process terminates when the addition of any of the remaining variables would yield a partial F-statistic greater than PIN (Jobson, 1992a). With categorical variables, the stepwise procedure choose the subset of explanatory variables by calculating the chi-square statistic $-2[\ln L_0 - \ln L_i]$ at each step with L_0 being the likelihood function when only the intercept is fitted whereas L_i refers to the likelihood function when at each step variables entered the model. At each stage, it selects the term giving the greatest improvement in fit (Agresti, 2002; Jobson, 1992b). Finally, for each model, the McFadden's Pseudo R^2 index was assessed to estimate how the model fits the data:

$$R^2 McFadden = 1 - \frac{LL_{full}}{LL_0} \tag{2}$$

where LL_{full} represents the loglikelihood of the model of interest which includes all statistically significant variables and LL_0 is the loglikelihood of the null model. The McFadden's Pseudo R^2 variability range is between 0 and 1; however, McFadden's Pseudo R^2 greater than 0.20 indicates a very good fit (Andreß et al., 2013).

For choosing the correct model, the Likelihood Ratio test (commonly known as LR test) is estimated as part of the random ordered/unordered model to determine the significance of the random formulation relative to the standard ordered/unordered logit model. The LR test compares the likelihood of the mixed model with the likelihood of the standard model:

$$\text{LR test} = -2\log\left(\frac{\text{L}_{\text{MIXED}}}{\text{L}_{\text{ST}}}\right) = -2(\text{LL}_{\text{MIXED}} - \text{LL}_{\text{ST}}) \tag{3}$$

where LL_{MIXED} is loglikelihood of the mixed model whereas LL_{ST} is the loglikelihood of the fixed parameters model. The likelihood ratio test statistic has an approximate χ^2distribution with k (the number of predictors) degrees of freedom. If the LR test p-value < 0.05 the random parameter logit is superior to the standard model with over 95% confidence. This indicates that the RPML model provide a statistically superior fit relative to the traditional fixed-parameter models (Anastasopoulos and Mannering, 2011; Seraneeprakarn et al., 2017).

All the explanatory variables reported in the Crash database paragraph have been transformed into dummy variables, through a complete disjunctive decoding process. Predictors with multiple categories (k) have been converted to a series of indicator variables (dummy variables) with k - 1 variables, the k-th dummy variable was not inserted in the model to avoid incurring in a problem of perfect multi-collinearity. All indicator variables were used to estimate the four logistic regression models and tested for inclusion. Each indicator variables variable was assessed for its importance to injury severity using the z-test with a significance level of 10%. All four econometric models were developed using the STATA software.

3.1.2 Multinomial logit

The Multinomial Logit is an "upgrade version" of binary logit regression as it tolerates two or more categories of the outcome variable. Crash severity analysis can be carried out considering the three classes (slight injury, serious injury and fatal crashes) as possible discrete outcomes (i.e., using British and Swedish databases) or just considering injury and fatal as the only two possible discrete outcome if using the Italian national database. In the first case, the model will be a multinomial logit whereas the latter condition gives rise to a binary logit model. Considering the general case of a multinomial logit model and more than two crash injury severity outcomes, the propensity of crash i (i=1,…,I) towards severity category j (j=1,…,J) is represented by severity propensity function (Washington et al., 2011):

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \beta_j' x_{ij} + \varepsilon_{ij} \tag{4}$$

where x_{ij} is a (K×1)–column vector of K exogenous attributes (geometric variables, environmental conditions, driver characteristics, …) that affect pedestrian injury severity level j, β_j is a (K×1)–column vector of the estimable parameters for crash severity category j, and ε_{ij} is the disturbance term assumed to be an independently and identically distributed (iid), following the Type I generalized extreme value distribution (i.e. Gumbel) with mean equal to zero, variance equal to one, and scale parameter η, as shown by McFadden (1981) and Washington et al. (2011).

For a standard multinomial logit, the utility is linear in β, then $V_{ij} = \beta_j x_{ij}$. Eachβ_j represents the estimated impact of variable x_{ij} on the response variable y_i. Increasing $\beta_j$indicates increasing severity whereas a negative β_j indicates decreasing severity. The standard multinomial logit (MNL) formulation takes the form:

$$P(y_i = j) = P_i(j) = \frac{e^{(\beta_j' x_{ij} + \varepsilon_{ij})}}{\sum_{j=1}^{J} e^{(\beta_j' x_{ij} + \varepsilon_{ij})}} \tag{5}$$

The parameter vectors (β's) are estimated by maximum likelihood and correspond to the effects of explanatory variables on outcome-specific level of severity. In a standard MNL formulation, β's are assumed fixed across observations and standard MNL is considered a fixed-parameter model meaning that the estimated coefficients represent the observation averaged effects without considering crash individual's diversity.

3.1.3 Random parameter multinomial logit

Random parameter models have been widely used in transportation research due to their flexible functional form compared to fixed-form model specifications. These methods tend to be preferred when, analysing the data, there is no strong a priori theoretical reason to prefer one functional form to the other.

The random parameter multinomial logit, also known as mixed multinomial logit model, is the generalized form of the multinomial logistic regression. In the fixed parameter logit model, estimated coefficients represent the averaged effects without considering crash individual's diversity. In the mixed model, instead, the coefficients of any of the variables are not limited to a fixed value but are allowed to vary across observations or analyst specified groups of observations. They are considered to be random and can be decomposed into their means b and deviations $\tilde{\beta}_j$(Mannering et al., 2016):

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \beta_j' x_{ij} + \varepsilon_{ij}; \beta_{\mathrm{j}} = b + \tilde{\beta}_j \quad (6)$$

where x_{ij} is a (K×1)–column vector of K exogenous attributes (geometric variables, environmental conditions, driver characteristics, …) specific to crash i that affects crash severity level j, β_j is a crash specific (K×1)- column vector of corresponding parameters that varies across crashes based on unobserved crash-specific attributes, and $\varepsilon_{ij}$is the disturbance term assumed to be iid across crash severity levels and crashes.

Hence, the standard multinomial logit hypotheses are relaxed (i.e., mixed logit does not exhibit independence from irrelevant alternatives) and one or more parameters can be randomly distributed in the mixed model. Indeed, the presence of correlation between unobserved characteristics of each individual will violate disturbance independence assumptions for error terms leading to erroneous parameter estimates whereas random-parameter model addresses unobserved heterogeneity with parameters that vary across individual observations: β_j' vector has a density function which is described by a vector of parameters θ (mean and variance). If unobserved heterogeneity is allowed, β_j is a vector with a continuous density function Prob ($\beta_j = \beta$) = $f\,(\beta|\theta)$, which means that the unconditional probability of individual i experiencing the severity level j from the set of severity outcomes J is obtained by considering the integrals of standard multinomial logit probabilities over a density of parameters and can be expressed in the form (McFadden, 1981; Train, 2009; Mannering et al., 2016):

$$P_i(j) = \int \frac{e^{\beta'_j x_{ij}}}{\sum_J e^{\beta'_j x_{ij}}} f\,(\beta|\theta) d\beta \quad (7)$$

The random multinomial logit probability is expressed as a weighted average of the probability evaluated with the multinomial logit formula at different values of β, with the weights given by the density function f(β). Standard multinomial logit is a special case of the mixed logit formulation because if β_j= b per each observation, there is no crash-specific unobserved heterogeneity among data and the random-parameter model degenerates at the standard multinomial logit with fixed parameter b (McFadden, 1981; Train, 2009; Mannering et al., 2016), and f(β)=1 for β_j= b while it is 0 for $\beta_j \neq$b.

3.1.4 Ordered logit

However, the MNL model disregarded the ordered nature of injury severity levels and treats them as independent alternatives. Thus the ordering information is lost (Srinivasan, 2002). With ordered outcomes, adjacent alternatives are expected to share some common trends depending on their proximity to each other - the closer they are, the bigger the trend they share. This potentially implies that adjacent response outcomes could also share some unobservable effects. In view of this fact, some of the standard unordered response models which are built on the assumption that unobserved effects are independent across alternatives, could provide inconsistent estimates when applied to ordered response outcomes. This suggests that considering a modeling framework that accounts for the ordinal nature of response outcomes is crucial when modeling the injury severity of traffic crashes (Abay, 2013).

The model is based on the cumulative probabilities of the response variable and it is assumed that the logit of each cumulative probability is a linear function of the covariates with regression coefficients constant across response categories. In this case, the effects of the explanatory variables on the severity levels are assumed to be fixed across observations. In other words, ordered logistic regression assumes that the coefficients that describe the relationship between the lowest versus all higher categories of the dependent variable (crash severity in our study) are the same as those that describe the relationship between the next lowest category and all higher categories, and so on. This is also called the proportional odds assumption, the parallel regression assumption or the grouped continuous model (Long, 1997). Assuming the severity of a crash as an ordered discrete variable with j categories (slight, serious, and fatal), three levels are given meaningful numeric values, usually 0*,* 1*, ..., J* (*J* upper limit). Slight, serious, and fatal might be labelled 0*,* 1*,* 2 and the numerical values represent a ranking so that, for crash severity, label “1” is more severe than “0” in a qualitative sense and the difference between “2” and “1” is not the same as that between “1” and “0”. In this case, although the numerical outcomes are merely labels of non-quantitative outcomes, the analysis will nonetheless have a regression-style motivation (Greene and Hensher, 2010). The severity propensity function is assumed as reported in equation 4 and the ordinal response y_i can be expressed as:

$$y_i = \begin{cases} 0\ if\ -\infty \le U_i \le \mu_1 \\ j\ if\ \mu_{j-1} < U_i \le \mu_j \\ J\ if\ \mu_{J-1} < U_i \le +\infty \end{cases} \tag{8}$$

where μ_j and μ_{j-1} represent the upper and lower thresholds for injury severity J *(value of cutoff or cut-points)*. The cumulative probability can be written as (Long, 1997):

$$P_i(j) = \frac{e^{(\beta'_j x_{ij} + \varepsilon_{ij} - \mu_j)}}{1 + e^{(\beta'_j x_{ij} + \varepsilon_{ij} - \mu_j)}}, j = 0,2, \ldots, J - 1 \tag{9}$$

The probability of outcome j corresponds to the area of the error distribution between the cut-points μ_{j-1} and μ_j. Φ represents the cumulative distribution function of the error logistic distribution with a mean of zero and a variance of $\pi^2/3$.

3.1.5 Random parameter ordered logit

The random parameters ordered logit model allows the thresholds in the ordered logit model to vary based on both observed as well as unobserved characteristics. It also accommodates unobserved heterogeneity in the effect of exogenous variables on injury propensity and the threshold values through a suitable specification of the thresholds relaxing the restriction of identical thresholds (Srinivasan, 2002). As for mixed multinomial logit, equation 11 determines the probability that crash i will result in injury-severity level j. Hence, both β's and threshold μ can systematically vary across crashes due to observed and unobserved factors: in an ordered random parameter logit, the thresholds also consist of a systematic component V_j and an unobserved disturbance error-terms τ, thus allowing for unobserved variability and randomness in the thresholds as expressed by the formula below:

$$\mu_{ij} = V_j + \tau_{ij} \tag{10}$$

Finally, the likelihood function for individual *i* represents the probability of injury severity actually experienced by that individual and can be evaluated as:

$$P_i(j) = \int \frac{e^{(\beta'_j x_{ij} + \varepsilon_{ij} - \mu_j)}}{1 + e^{(\beta'_j x_{ij} + \varepsilon_{ij} - \mu_j)}} f\,(\beta|\theta) d\beta\ j = 1,2, \ldots, J - 1 \tag{11}$$

Therefore, to account for these circumstances, a random parameter ordered logit model was developed to capture the unobserved heterogeneity, which is achieved by adding a randomly distributed error term.

3.2 Machine Learning Models

In this research five popular machine learning tools were developed and compared to each other and with the econometric models. The algorithms implemented included: 1) classification tree, 2) random forest, 3) association rule discovery, 4) artificial neural network, and 5) support vector machine. Below, each method was presented and described. All the models' hyper parameter values were tuned for optimized results using grid search technique. The reason why the machine learning algorithms have been of keen interest to researchers and analysts is due to their ability to work with a very high number of variables and choose, among them, the best subset of predictors containing features that are highly correlated with the response class, yet uncorrelated with each other. Furthermore, the process of feature subset selection identifies and removes as much irrelevant and redundant information as possible in order to reduce the dimensionality of the data and allow learning algorithms to operate faster and more effectively (Maglogiannis et al., 2007). It does automatically avoid the correlation issue.

3.2.1 Classification tree

A CT is a non-linear and non-parametric tool and an oriented graph where the root node is divided into leaf nodes by an explanatory variable also called a splitter. All independent variables are a candidates for the splits at each internal node of the tree. However, only the predictor giving the best partition is chosen. In our study, we developed the CART algorithm introduced by Breiman et al. (1984) and the impurity at each node was assessed by the Gini reduction criterion (Higher the value of Gini index, higher the homogeneity of the node due to the split) which can be calculated as follows:

$$i_Y(t) = 1 - \sum_j p(j|t)^2 \tag{12}$$

where P(j|t) is the proportion of observations in the node t that belong to the class j. If a node is 'pure', all the observations in the node belong to one class and the impurity of that node is zero.

The total impurity of any tree T is defined as follows:

$$i_Y(T) = \sum_{t \in \tilde{T}} i_Y(t)p(t) \tag{13}$$

where $i_Y(t)$ is the impurity of the node t, p(t) = N(t)/N is the weight of the node t, N(t) is the number f observations falling in node t, N is the total number of observations, and $\tilde{T}$ is the set of terminal odes of the tree T. By definition, the terminal nodes present a low degree of impurity compared with the root node.

The total impurity of the tree is reduced by finding at each node of the tree the best partition of the observations into disjoint classes which are externally heterogeneous and internally homogeneous. Tree growing was stopped basing on two criteria: (1) the reduction in the Gini measures is less than a prespecified minimum fixed equal to 0.0001 (default value); and (2) the maximum number of levels of the tree equal to 4.

The class assigned to each node was selected according to the greatest value of the posterior classification ratio (PCR) evaluated for that node. The PCR, introduced by Montella et al. (2011), compares the classification of the terminal nodes of the tree with the classification of the root node and is calculated as follows:

$$\mathrm{PCR}\ (\mathrm{j}|\mathrm{t}) = \frac{p\ (j|t)}{p\ (j|t_{root})} \tag{14}$$

where p(j|t) is the proportion of observations in node t that belong to the class j and t_{root} is the root node of the tree.

Among the outputs of the CART algorithm, the tool provides the variable importance (also known as predictor ranking) based on the contribution of the predictors in building up the tree. It can be obtained by the Variable Importance Index (VIM), which reflects the impact of the predictor variables on the model. The information is obtained for all the independent variables, making it easy to find which ones are the most important. Therefore, the relative importance of a variable x_j is defined in the following equation (Kashani and Mohaymany, 2011):

$$\mathrm{VIM}\ (\mathrm{x_j}) = \sum_{t=1}^{T} \frac{n_t}{N}\ \Delta\mathrm{Gini}\ (\mathrm{x_j}, \mathrm{t}) \tag{15}$$

where ΔGini (x_j , t) is the Gini reduction at a node t that is achieved by splitting by the variable x_j, $\frac{n_t}{N}$ is the proportion of the observations in the dataset that belong to node t, T is the total number of nodes, and N is the total number of observations.

CT was carried out with SPSS software.

3.2.2 Random forest

CT, despite its advantages, is sometimes found to generate unstable predictions given certain perturbations (Breiman, 1996). To improve stability, Breiman (2001) proposed the RF method, which constructs an ensemble of B trees $\{T_1(X), ..., T_B(X)\}$, where $X_i = \{x_{i1}, ..., x_{ip}\}$ is a p-dimensional vector of descriptors or properties associated with a crash. The multiple identically distributed CTs in the forest are independently constructed by considering a random subset of attributes and having access to a random set of data N_t. The ensemble produces B outputs $\{\hat{Y}_1 = T_1(X), ..., \hat{Y}_B = T_B(X)\}$ where $\hat{Y}_b$, b = 1, ..., B, is the prediction for a crash by the b^{th} tree. Outputs of all trees are aggregated to produce one final prediction, $\hat{Y}$. In classification process, RF lets each tree vote for the predicted class and uses the class having the majority votes as the final output of the prediction process. RF combines the generated decision trees to minimize the model bias and variance. To further avoid the overfitting, the hyperparameters were turned, such as the tree maximum depth (set equal to 4). For each tree, out of bag data (OOB, N - N_t observations) were used to assess the misclassification error which evaluates the optimum number of trees making up the "forest" and to estimate the importance of the variables. Given data on a set of n crashes, D = {(X1, Y1), ..., (Xn, Yn)}, where Xi is a vector of descriptors and Yi is either the corresponding class label for the i^{th} crash with i = 1, ..., n, the algorithm proceeds as follows:

(1) bootstrap sample, create a random sample with replacement from the original sample with sample size N_t replicated B times;

(2) for each bootstrap sample, grow the tree using CART algorithm choosing at each node the best split among a randomly selected subset of descriptors;

(3) repeat the above steps until B trees are generated.

However, it has been shown that there is a potential overestimate of the true prediction error depending on the choices of the random forests hyperparameters, such as the number of trees (B) and the number of descriptors. To reduce the true prediction error, the Out-of-Bag Estimate of the error rate (ER^{OOB}) was estimated varying B and the number of descriptors:

$$\mathrm{ER}^{\mathrm{OOB}} = \frac{\sum_{i=1}^{N}\left(\hat{Y}^{OOB}(X_i) \neq Y_i\right)}{N} \tag{16}$$

where $\hat{Y}^{OOB}(X_i)$ is the predicted class for the i^{th} observation, X_i is the vector of the attributes of the i^{th} observation, Y_i is the class label of the i^{th} observation, and N is the total number of observations. The values of B and mtry have been chosen in correspondence of a stable $\mathrm{ER}^{\mathrm{OOB}}$ around the minimum value. During the classification process, the OOB data are left out from the training trees and are then utilized to achieve unbiased estimate of variable importance as trees are added to the forest. For each growing tree in the forest, the prediction error rate on the OOB data is recorded and the procedure is done after permuting each predictor variable. The differences between the two error rates would be averaged over all grown trees and then normalized by the standard deviation of the differences. The averaged differences would be the raw importance score for the variables. In this research, random forest model was developed to perform variable importance ranking, which would allow to understand the most important variables in the crash severity analysis according to the tool. The variable importance ranking is measured by the classification accuracy and Gini impurity coefficient. Further outputs of the model were the trees which can be extracted. The tool reiterated until it does not find the number of tree and the number of variables which provide the highest prediction accuracy.

The variable importance measure for variable x_j is computed as the sum of the importances over all trees in the forest:

$$VIM(\mathrm{x}_j) = \frac{\sum_{t=1}^{ntree} VIM^t(\mathrm{x}_j)}{ntree} \tag{17}$$

where $VIM(\mathrm{x}_j)$ is the variable importance of the t^{th} tree calculated using equation (15) and ntree is the number of trees.

RF was performed in the R-cran software environment using the package “randomForestSRC”.

3.2.3 Association rules

Association rule discovery is a descriptive analytic method for extracting and refining valuable knowledge from large datasets. The tool belongs to data mining techniques and has already shown its ability in discovering significant rules highlighting items that occur frequently together in a crash dataset. What is more, association rule is focused on the search and finding of patterns in data rather than the confirmation of hypotheses (Das et al., 2019) and is not affected by the absence of important data, which can potentially undermine traditional statistical analyses leading to biased and inconsistent results and erroneous safety engineering countermeasures. Association discovery was performed using the a priori algorithm (Agrawal et al., 1993). Each crash record contains different items (e.g., crash type, crash severity, alignment, grade, pavement conditions, etc.) and the dataset contains all the items of each crash. Basing on the relative frequency of times the item-sets occur alone and in combination in a dataset, the association rules were extracted with the form “A → B”, where A and B are disjoint item-sets: A is the antecedent and B is the consequent. The a priori algorithm uses simple and repetitive steps examining candidate item-sets to find frequent item-sets. Then, it uses the new candidate item-sets to find new frequent item-sets until no newer item-sets can be produced (Montella et al., 2020, 2021). The parameters support, confidence, and lift were used to assess the strength of each association rule. Support is the percentage of the entire data set covered by the rule, confidence measures the reliability of the inference of a generated rule, and lift is a measure of the statistical interdependence of the rule.

Supports are calculated as follows:

$$\text{Support}\ (\text{A} \rightarrow \text{B}) = \frac{\#(\,A \cap B)}{N};\text{Support}\ (\text{A}) = \frac{\#(\,A\,)}{N};\ \text{Support}\ (\text{B}) = \frac{\#(\,B\,)}{N} \tag{18}$$

where support(A ⟶ B) is the support of the association rule, support(A) is the support of the antecedent, support(B) is the support of the consequent, #(A ∩ B) is the number of crashes where both the condition A (antecedent) and the condition B (consequent) occur, #(A) is the number of crashes with A antecedent, #(B) is the number of crashes with B consequent, and N is the total number of crashes in the dataset.

Confidence is calculated as follows:

$$\text{Confidence} = \frac{\text{Support(A→B)}}{\text{Support (A)}} \tag{19}$$

Confidence is defined by the percentage of cases in which a consequent appears given that the antecedent has occurred. A high confidence for A → B indicates that the presence of B as consequent is high in the crashes having A (single item or combination of more items) as antecedent.

Lift is calculated as follows:

$$\text{Lift} = \frac{\text{Support(A→B)}}{\text{Support (A)} \times \text{Support (B)}} \tag{20}$$

The lift of the rule relates the frequency of co-occurrence of the antecedent and the consequent to the expected frequency of co-occurrence under the assumption of conditional independence. A lift value lower than 1 indicates negative interdependence between the antecedent and the consequent. A lift value equal to 1 designates independence, and a value greater than 1 indicates positive interdependence (i.e., the number of times the sets of items occur together is greater than they would if they were independent of each other). The higher the lift, the greater the strength and the interest of the association rule since it would indicate how more often the antecedent and the consequent are part of the same crash than if these events were statistically independent. It is desirable for the rules to have a high level of support, a large confidence, and a lift value considerably greater than one. Thus, minimum values for support, confidence and lift are needed.

A rule with a single antecedent and a single consequent is defined as a 2-item rule; similarly, a rule with two antecedents and single consequent is defined as a 3-item rule. Each rule with n+1 items is validated by verifying that each variable produces a lift increase (LIC). The LIC ensures that each additional item in the rules leads to an increase in terms of lift. The rules with only one item in the antecedent are used as a starting point, rules with more items are selected over simpler rules if the LIC condition satisfied the minimum threshold of 1.05 (López et al., 2014; Montella et al., 2011, 2020, 2021).

LIC is calculated as follows:

$$\mathrm{LIC} = \frac{\mathrm{Lift}_{\mathrm{A_n}}}{\mathrm{Lift}_{\mathrm{A_{n-1}}}} \tag{21}$$

where A_{n-1} is the antecedent of rule with n-1 items, and A_n is the antecedent of rule with n items.

These criteria are further explained with an example. Let's suppose that a crash database consists of 1,000 crashes and out of these crashes 100 were fatal. Out of the total crashes, 400 of them occurred on curve alignment and 80 of them were fatal. Now consider the rule "curve alignment → fatal crashes" for this database. In this rule, "curve alignment" is the antecedent while "fatal crashes" is the consequent. The support for the rule is defined as the percentage of all crashes that were both fatal and occurred on a curve alignment. For the aforementioned hypothetical rule, support would be 8% (80/1,000 = 0.08). Confidence for the rule is defined as the percentage of fatal crashes among all crashes that occurred on curve alignment. The number of such crashes is 400 and hence in this database, the confidence for the aforementioned rule would be 20% (80/400 = 0.20). The lift is the ratio between the support of the rule (equal to 80/1000 = 0.08) and the expected support under the assumption of conditional independence (support of curve alignment, equal to 400/1,000 = 0.40, multiplied for the support of fatal crashes, equal to 100/1,000 = 0.10; 0.40 × 0.10 = 0.04) and is equal to 2 (0.08/0.04). Now, let's make a new assumption. Out of the 400 crashes occurred on curves, in 100 crashes drivers were speeding and 30 of these crashes were fatal. The new rule is "curve alignment & speeding → fatal crashes". This is a 3-item rule with support equal to 0.03 (30/1000 = 0.03), confidence equal to 0.30 (30/100 = 0.30), and lift equal to 3 (0.03/(0.10 × 0.10) = 0.03/0.01 = 3). The parent rule (curve alignment → fatal crashes) has a lift equal to 2 and the lift increase of the 3-item rule (curve alignment & speeding → fatal crashes) is 1.5 (3/2 = 1.5). It means that the proportion of fatal crashes for "curve alignment & speeding" is 1.5 times the proportion for "curve alignment".

The threshold values of support and confidence depend on the dataset characteristics. Conversely, as the lift is used to assess the dependence between the items in the item set, the threshold value depends on how much stronger the analyst wants the dependence

AR was performed in the R-cran software environment using the package "arules".

3.2.4 Support vector machine

Support Vector Machine (SVM), developed by Cortes and Vapnik (1995), is an emerging machine learning technique in statistical learning theory of multi-dimensional function which is used for classification and regression analysis. It holds the ability to be universal approximator of any multivariate functions to any desired level of accuracy. According to the previous studies (Wei et al., 2013), SVM has been used in different engineering fields with good accuracy. SVM is used to develop an optimal separating hyperplane to categorize the observations into several groups while maximizing the margin between the decision boundaries and minimizing the empirical error. In our study, each crash represent a set of points in N-dimensional space and SVM generates a (N–1) dimensional hyperplane to split those points into groups. The distance of the closest points to this hyperplane on each side is maximized. Hence, the plane constitutes the decision boundaries and the hyperplane is a p − 1 dimensional plane. The hyperplane has the following equation:

$$y(x) = w^T x + b = 0 \tag{22}$$

where hyperplane $y(x) = 0$ defines a decision boundary in the N-dimensional space, w represents the parameters of a vector perpendicular to the hyperplane and b is the bias. The normal vector and bias are determined through the learning procedure on a training set which includes the full set (x_n, y_n) of crash-related explanatory variables x_n while y_n represents the injury severity outcomes. The construction of the higher dimensional space is based upon the concept of a kernel function. The readers may refer to Kecman (2005) for the basic understanding of the working principle of SVM. Although several kernel functions exist, Radial Basis Function (RBF) is the most commonly used for crash severity analyses since it is capable of capturing the non-linearity relationships between crash severity and explanatory variables (Assi et al., 2020; Li et al., 2012; Mokhtarimousavi et al., 2019;Zhang et al., 2018, Yu and Abdel-Aty, 2013). The decision boundaries may or may not be linear depending on the pre-set kernel function. Radial basis function (RBF) is the most commonly used for crash severity analyses since it is capable of capturing the non-linearity relationships between crash severity and explanatory variables:

$$K(X_i, X_j) = \exp(-\Upsilon \, |X_i - X_j|2), \Upsilon > 0 \tag{23}$$

where:

X_i and X_j are vectors of explanatory variables for the i[th] and the j[th] crashes;

$|X_i - X_j|^2$ is the euclidean distance between two crashes X_i and X_j;

$\Upsilon = 1/\sigma^2$ where σ^2 is the variance of samples selected by the model as support vectors.

The development of the SVM model also depends on the penalty parameter C of the error term. It controls the trade-off between smooth decision boundaries and classifying the points correctly, and it is calculated as follows:

$$\mathrm{ER}^{\mathrm{SVM}} = \frac{\sum_{i=1}^{N}(\hat{Y}^{SVM}(X_i) \neq Y_i)}{N} \quad (24)$$

where:

$\hat{Y}^{SVM}(X_i)$ is the predicted class for the i[th] crash;

X_i is the vector of descriptors of the i[th] crash;

Y_i is the lass label of the i[th] crash;

N is the total number of crashes.

To determine the separability of the optimal hyperplane, grid search was used for the joint optimization of C and γ parameters and feature selection. This approach methodically builds and evaluates a model for each combination of algorithm parameters (γ and C) specified in a grid. For each model, the classification error was used as a performance measure. The combination of hyper-parameters with the lower classification error was chosen to develop the optimal hyperplane.

The variables contributing to the separability of the optimal hyperplane provide an indication of the relative importance of the variables to the separation. Theoretically, it has less overfitting problem and better generalization ability. However, the main problem in constructing the SVM model is to adequately select training parameter values as an inappropriate parameter setting leads to poor prediction accuracy.

SVM was performed in the R-cran software environment using the packages “caret” and “e1071” where the function svm(·) includes C- classification problems and RBF kernel function.

3.2.5 Artificial neural networks

As CT and RF, also ANN is an oriented graph inspired by a biological neural network. Like the structure of the human brain, the ANN models consist of neurons in a complex and non-linear form. The ANN models work by creating a non-linear relationship between dependent and independent variables depending on a set of experimental data. The neurons are connected to each other by weighted links. ANNs consist of a layer of input nodes and a layer of output nodes connected by one or more layers of hidden nodes. The input layer nodes pass information to the hidden layer nodes by firing activation functions, and the hidden layer nodes fire or remain dormant depending on the evidence presented. The hidden layers apply weighting functions to the evidence, and when the value of a particular node or set of nodes in the hidden layer reaches some threshold, a value is passed to one or more nodes in the output layer.

The technique creates a feed-forward multilayer perceptron ANN that consists of multiple nodes (or neurons) organized in three or more layers with a back-propagation learning process to minimize classification errors. In this research, a three-layer network has been implemented, as previous studies suggest that ANNs with a singular hidden layer are less likely to be trapped at a local minimum (de Villiers and Barnard, 1993; Zeng et al., 2016). Thus, the information flows from the input layer, passes through the hidden layer, and then to the output layer to produce a classification. The hidden layer has $1 + \sum_{p=1}^{P} k_p$ neurons (let's consider a dataset containing P independent variables classified on k_p potential risk factors having effects on crash severity), each risk factor is represented by a node while another constant node is included representing the bias. The output layer has three neurons according to the three severity levels in the study. The ANN tool operates in two distinct phases: the training phase and the testing phase. Firstly, the learning algorithm is used to train the network, then a testing algorithm is used for testing the network.

One of the most important components of Artificial Neural Networks is the Loss Function. Loss is nothing but a prediction error of the Artificial Neural Net assessed through the Loss Function method. For multi-class classification tasks, Cross-entropy is the loss function most commonly used. The use of cross-entropy implies that there must be the same number of output nodes as the classes of the target variable. The final layer output should be passed through a softmax activation function so that each node output has a probability value ranging between 0 and 1. Basically, the target vector would be of the same size as the number of classes and the index position corresponding to the actual class would be 1 and all others would be zero.

Thus, the neurons of the input layer transfer information to the hidden layer through the hyperbolic tangent activation function and, from the hidden layer to the output layer, through the softmax function used for classification purposes:

$$\mathrm{z} = \mathrm{softmax}[\sum_{\mathrm{j}=1}^{\mathrm{J}} \mathrm{w}_{\mathrm{j}}^{(2)} \tanh(\sum_{\mathrm{p}=1}^{P} \mathrm{w}_{\mathrm{j,p}}^{(1)} \mathrm{k}_{\mathrm{p}})] \tag{25}$$

Assuming J being the number of neurons in the hidden layer, $w_{j,p}^{(1)}$ is the connection weight between hidden node j, j=1, …, J and input node p, p=1,…, P whereas k_p are the factors. In the output layer, Z nodes expresses severity outcomes predicted by ANN (Z=3 in the British and Swedish case studies, Z=2 in the Italian case study) and y_i is the ith observed response in the dataset. If for the ith crash, y_i = z, then z=1 while z=0 otherwise. $\mathrm{w}_{\mathrm{j}}^{(2)}$ is the weight of the connection between output node z and hidden node j.

The connection weights were estimated using a back-propagation learning process to minimize classification errors. Standard back-propagation is a gradient descent algorithm in which the network weights are moved along the negative of the gradient of the performance function. The combination of weights that minimizes the error function is considered a solution to the learning problem. The backpropagation algorithm proceeds as follows:

(1) The back-propagation algorithm starts with random weights, and the goal is to adjust them to reduce this error until the ANN learns the training data;

(2) If the expected output is not obtained, backward propagation begins. The difference between the actual and the expected output is calculated recursively and step by step and the error is returned through the original link access;

(3) The weight and the value of each neuron are then modified and transmitted successively to the input layer, and the forward multilayer perceptron restarts.

These two processes of forward multilayer perceptron and back-propagation error are repeated so that the error gradually decreases. The goal is to minimize the error by adjusting the weights so that optimum weights are obtained after the error backpropagation.

The gradient (G) of a weighting to the error, total error (E) and total mean square errors (e_p) are defined as:

$$G = \frac{\partial E}{\partial w} \tag{26}$$

$$E = \sum e_p \tag{27}$$

$$e_p = \frac{1}{2}\sum_{k-1}^{m}\left(y_k^p - \bar{y}_k^p\right)^2 \tag{28}$$

where:

w is one of the network weightings w_{pl}, w_{jp}, w_{kj}; y_k^p is the actual output; and $\bar{y}_k^p$ is the expected output.

The adjustment of weight is calculated as:

$$\Delta w^{new} = -\eta G + \alpha \Delta w^{old} \tag{29}$$

where:

Δw^{new} is the present adjustment for weighting or for threshold; Δw^{old} is the immediate past value of its counterpart; α is a dynamic coefficient and it takes value in the range between 0 and 1; and G is the gradient of a weighting to the error.

This procedure was applied to categorical data after transforming the categorical variables into dummy variables, through a complete disjunctive decoding process. Predictors with multiple categories (k) have been converted to a series of indicator variables (dummy variables) with k variables.

The importance of a specific explanatory variable is determined by identifying all weighted connections between the nodes of interest. All weights connecting the specific input node that passes through the hidden layer to the specific response variable are identified. This is repeated for all other explanatory variables until all weights that are specific to each input variable are determined.

ANN was performed with the SPSS software.

3.3 Dealing with imbalanced data

In our datasets, fatal and serious crashes represent a small share of the total crashes with order ratios of 2:100 for fatal crashes and 25:100 for serious injury crashes in Great Britain, 2:100 for fatal crashes and 4:100 for serious injury crashes in Sweden, and 3:100 for fatal crashes in Italy. This means that the response variable distributions are extremely imbalanced (Chawala et al., 2002; Chawala et al., 2004; Fiorentini & Losa, 2020). In this case, the learning process can lead to distorted results (He & Garcia, 2009; Ndour et al., 2012) as the model has adequate information about the majority class (i.e., slight injury) but insufficient information about your minority class (fatal and serious injury crashes). That is why there will be high misclassification errors for the minority class. From the analysis of the literature review, several techniques to take the imbalanced data issue into account have been proposed overtime. Sampling techniques are the most traditional choice when dealing with the problem of imbalanced classes. Sampling techniques implies both oversampling and undersampling. Oversampling replicates instances from the minor class and repeats them until all the classes have equal frequency whereas undersampling discards the majority class instances until the majority class reaches the size of the minor classes. It only takes into account the closeness of the data whereas the intrinsic characteristics are not taken into consideration (Sáez et al., 2015). Furthermore, both sampling techniques implies changes in the original dataset. As a matter of fact, it is well known that the main limitation of these techniques includes creation of distorted samples around the decision boundary of the majority and minority classes thereby disrupting the natural boundary between classes. However, the techniques do not necessarily improve the minority class performance (Islam et al., 2021). Another group of techniques implies the use of costs in the pre-classification or post-classification process. There are different ways eligible to account for the misclassification costs. The post-classification approach, also known as cost-sensitive technique, operates after the tools has provided its results. The purpose is to penalize the misclassification by setting a higher cost for misclassified fatal and serious crashes and, at the same time, reducing the misclassification cost for the slight injury. So that a small cost, given to the majority class, results in a small penalty and a small update to the model coefficients whereas large cost, given to fatal and serious injury crashes, result in a large penalty and a large update to the model coefficients. The pre-classification process operates in the early stage of the classification process, before the algorithm classification modifying the way the algorithm account for the skewed distribution of the classes and giving different weights to both the majority and minority classes. The difference in weights will influence the classification of the classes with the aim of penalizing the misclassification made by the minority class by setting a higher class weight and at the same time reducing weight for the majority class.

To solve the imbalanced distribution problem avoiding the limitations of the sampling techniques, in this research the weighted approach was used to improve the classifier performance. The weight class is automatically defined based on inversely adjusting weights proportional to class frequencies

Each cost (also called weight) can be assessed as follows (Singh, 2020):

$$\mathrm{w_k} = \frac{N_{crashes}}{n_c \times N_k} \tag{30}$$

where k = 1, 2, 3 (1 = slight injury, 2 = serious injury, 3 = fatal) for Great Britain and Sweden and k = 1, 2 (1 = injury, 2 = fatal) for Italy, w_k is the weight to assign to the respective level of severity k, $N_{crashes}$ is the total number of observations in the dataset, N_i is the number of crashes with the severity level “k”, and n_c = is equal to the number of crash severity levels considered in the data (3 for Great Britain and Sweden, 2 for Italy). Various empirical studies have shown that the learning which improves using costs or weights is superior to sampling methods. The main advantage of adopting the former approach over under-sampling and over-sampling is that the first did not change class distributions in the dataset to balance them but operates manipulating classifier algorithms internally.

3.4 Measures of performance

The aim of a classifier is to minimize the false positive rates (representing Type I error) and false negative rates (representing Type II error), maximizing the true negative and positive rates. The true negatives and positives as well as the false negatives and positives can be observed in the confusion matrix. The confusion matrix (or contingency table) has the predicted class instances on the columns, the rows denote the actual class instances, and the diagonal represents the accurate prediction. Starting from the simplest measures TP, TN, FN, and FP: TP measures the number of positives correctly identified (e.g., the number of severe injury and fatal crashes correctly identified) and TN denotes the number of negative examples that are classified correctly (e.g., the number of non-severe Injury and non-fatal crashes which are correctly identified as slight injury crashes), while FN defines the number of the estimated instances incorrectly classified as negative (e.g., the number of Severe Injury or Fatal crashes which are incorrectly identified as Slight Injury crashes) and FP denotes the number of misclassified negative examples (e.g., the number of slight injury crashes which are incorrectly identified as severe injury or fatal). For binary classification problems, the confusion matrix is as reported in Table 27. Both false negative and positive cases represent errors in the classification process. However, in crash severity analyses, a false negative has the most serious consequences than a false positive as it implies that a fatal or serious injury crash is misclassified as a slight injury crash.

Table 27 – A confusion matrix for binary classification.

		Predicted	
		Negative	Positive
Observed	Negative	**TN**	**FP**
	Positive	**FN**	**TP**

Nevertheless, to provide a wider framework of the ability of a classifier, multi-parameter indicators were preferred in lieu of the true positive/negative and false positive/negative rates.

Among the common performance metrics used to evaluate classification performance, accuracy and error rate are the most widely used. The first represents the percentage of correctly classified instances (corresponding to the sum of the diagonal elements in the confusion matrix) and divided by the total number of instances. The error rate, instead, corresponds to the sum of off-diagonal elements in the confusion matrix divided by the total number of instances and is the percentage of incorrectly classified instances. However, even though accuracy is the most commonly used indicator of model performance, in some specific situations, when the distribution of the response variable in the sample data is extremely imbalanced, like in presence of fatal and serious crashes usually occur less than slight crashes, accuracy has certain limitations. That means the accuracy is not a perfect performance metric. The error rate suffers from similar drawbacks. Firstly, it is easy to obtain high accuracy (or low error rate) under highly imbalanced problems. Secondly, these classifiers assume that errors are equally cost which is not true for imbalanced data where misclassifying instances of the minority class is generally much costlier than misclassifying instances of the majority class (Damji et al., 2020; Fernandez et al., 2018). As far as road safety is concerned, a correct classification of factors contributing to fatal crashes is a far cry from the correct identification of the factors contributing to slight injury crashes.

Hence, a set of metrics assessed in this research was chosen and reported below (Guo et al., 2008):

$$TN_{rate} = \mathrm{Acc}^{-} = \frac{\mathrm{TN}}{\mathrm{TN+\ FP}} = specificity \tag{31}$$

where Acc^{-} is the true negative rate, also known as specificity, TN is the number of true negatives, and FP is the number of false positives;

$$TP_{rate} = \mathrm{Acc}^{+} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} = Recall = sensitivity \tag{32}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{33}$$

where Acc^{+} is the true positive rate, also known as Recall or sensitivity. Precision measures the exactness of a classification algorithm (a low precision indicates many FPs). Recall measures the completeness of a classifier (a low recall means several FNs). However, recall and precision are often in tension with each other, as precision increases in a model, recall often drops. Thus, it is necessary to have a balanced classification threshold to manage the results of the two metrics. To overcome the issue, other evaluation metrics for multiclass classifiers with imbalance problems can be used, such as F-measure and G-mean. These two metrics were used to provide one or more parameters with one measure offering more comprehensive analysis assessments. F-measure (eq. 34) is the weighted harmonic mean of precision and recall (both referred to the minority class) and high F-measure usually indicates the model's good overall performance. G-mean (eq. 35) is the geometric mean between the percentage of negative examples and the percentage of positive examples both correctly recognized. F-measure and G-mean are commonly expressed as:

$$\text{F-measure} = \frac{(1+\beta^2)\times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision+Recall}} \tag{34}$$

$$\text{G-mean} = (\mathrm{Acc}^{-} \times \mathrm{Acc}^{+})^{\frac{1}{2}} \tag{35}$$

where β is a coefficient to adjust the relative importance of Precision versus Recall, set equal to 1.

AUC is the area under the receiving operating curve (ROC), a widely used graphical plot that illustrates the ability of a classifier created by plotting the true positive rate (TPR also known as sensitivity) on the vertical axis against the false positive rate (FPR also known as 1- specificity) on the horizontal axis at various threshold settings. When a ROC curve is created, AUC can be assessed. It indicates how capable the model is in distinguishing between classes. The curve that is closer to the upper left corner, its corresponding model has a better classification prediction ability. In other words, an AUC value varies between 0 and 1. However, a value of 0.5 for AUC indicates that the ROC curve will fall on the diagonal (i.e., 45-degree line) and hence suggests that the model has no classification ability. AUC greater than 0.60 is considered satisfactory (Kashani et al., 2014), AUC equal to 1 represents perfect classification.

Once evaluated the performance metrics for each class, the final values are the weighted mean of them, in which the relative frequencies of the classes on the data are their weights (Bina et al., 2013).

CHAPTER IV ~ STUDY DATA

This chapter provides a description of the three national databases used in this research. Firstly, for each database, it was reported the structure of the crash form and an overview of the crash information collected. Then, data related to the study period investigated in this dissertation were summarized and descriptive statistics of data were provided.

4.1 Great Britain data

Detailed road safety data about the circumstances of personal injury road crashes in GB has been collected in STATS19 dataset and provided by English Department of Transport (www.gov.uk) since 1979. Crash information is collected by police at the scene of a crash or reported by a member of the public at a police station. All reported crashes occurred on a public highway (including footways) in which at least one vehicle or a vehicle in collision with a pedestrian is involved and which becomes known to the police within 30 days of its occurrence whereas all crashes occurred on private land (including private drives) or car parks as well as damage only crashes that do not result in personal injury were not reported in the data. Crash information is then recorded using the STATS19 crash reporting form and available at https://www.gov.uk/transport-statistics-notes-and-guidance-road-accident-and-safety. The Stats19 form is really divided in more sections:

- A crash record form setting out the attendant circumstances associated with each crash. Information collected in these form includes: The road class, the road type, the speed limit, lighting conditions, weather and road surface conditions, presence or otherwise of junctions and pedestrian crossing facilities. Further information related to the date, time and location is also provided (Figure 6).
- A vehicle record: for each of the vehicles involved in the crash, a separate form is filled in providing details about the vehicle, its movements before and in the course of the crash, as well as its position. In this section data about driver (their age, gender, and journey purpose) are also collected and whether or not it was a hit and run crash Figure 7).
- A casualty record: as for each vehicle involved in the crash, also for each casualty there is a form to fill in. Information about the age and gender of the casualty and the severity of their injuries (fatal, serious or slight) is collected as well as details about the location and movements of any pedestrian casualties. There is also the presence of a casualty class that provides whether they

were a driver/rider, vehicle passenger (including car and bus passengers recorded separately), or a pedestrian (Figure 8).

- A contributory factors form. One of these forms is completed for each accident. The form contains a grid with 76 factors and up to 6 may be selected as contributing to a crash by the police officer if considered relevant to the crash. Each of these factors is linked to one of the person involved in the crash. The police officer also indicates whether the factor was 'very likely' to have contributed to the crash or only have a 'possible' link to it (Figure 9).

Originally, crash data were provided in three subsets reporting crash, vehicle, and casualty-related information reflecting the crash report forms provided from Figure 6 to Figure 8: (1) a dataset containing 32 different fields describing general and specific characteristics of the crash focusing the information on crash localization (latitude and longitude), road class and type, general characteristics such as weather, lighting, road surface condition, crash time, and specific information about the site when crash occurred such as presence of hazards on carriageway, special road surface conditions. Furthermore, crash data contained information on speed limit and functional road classification; (2) a dataset containing 22 variables describing crash involved vehicles focusing on vehicle type, presence of articulated vehicle or not, its age, engine (CC) and propulsion type, vehicle manoeuvres and an eventual object hit in/off carriageway, and driver information such as age, gender, and journey purpose; and (3) a dataset containing 16 variables describing all casualties (driver, passenger or pedestrian) involved in the crash. To obtain a unique set of information, the three subsets were merged into one by the use of crash index, which is unique for each crash. In this research, pedestrian crash data related to the three year period 2016-2018 were used (Rella Riccardi et al., 2022).

MG NSRF/A

ACCIDENT STATISTICS

***FATAL / SERIOUS / SLIGHT**

Incident URN

Other ref.

1.3 ACCIDENT REFERENCE

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

Figure 6 – STATS19 crash report form, crash general info section.

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

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

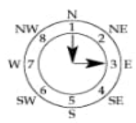
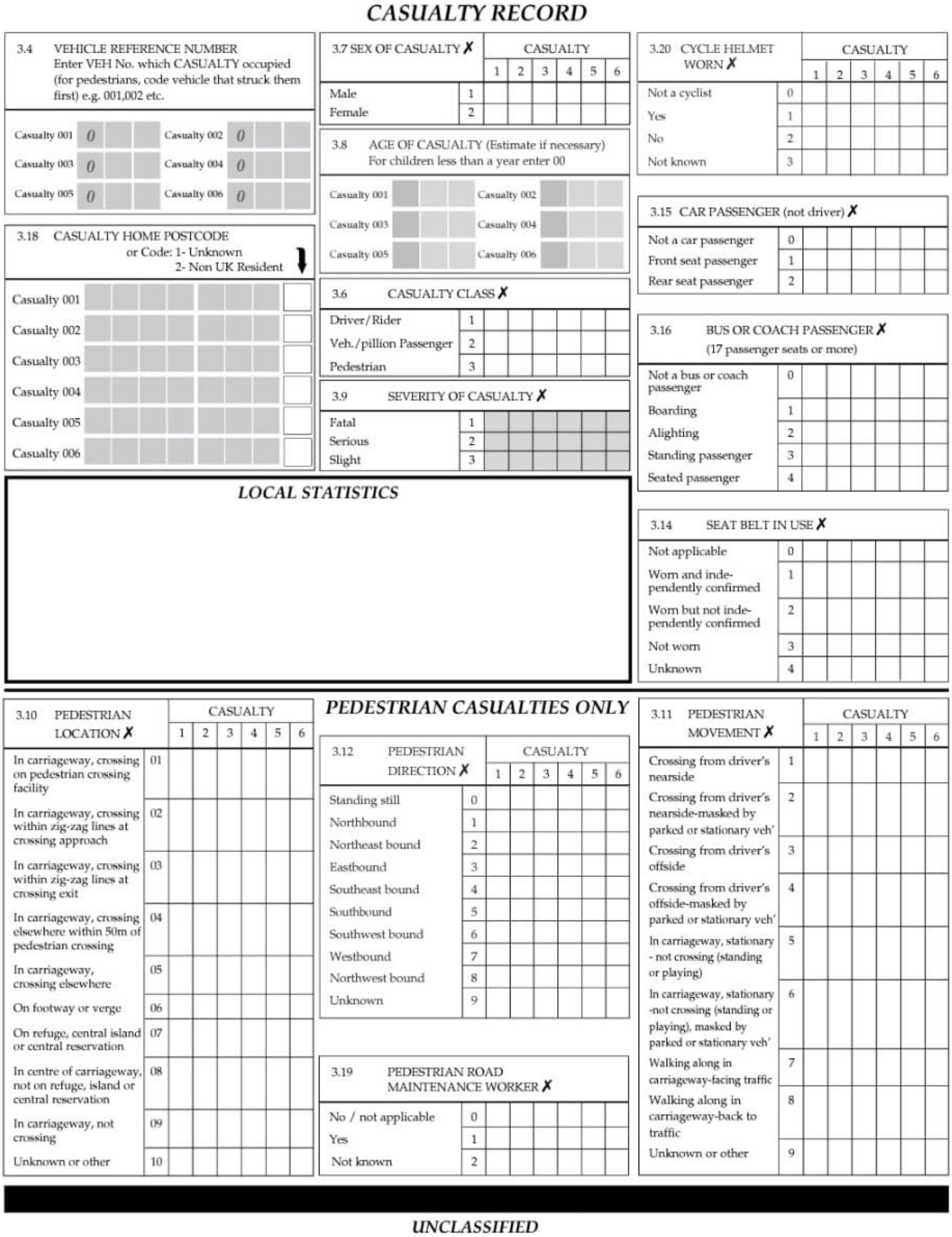


Figure 7 – STATS19 crash report form, vehicle data section.



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						

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						

LOCAL STATISTICS

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

Figure 8 – STATS19 crash report form, casualty data section.

Even though these factors reflect the reporting officer's opinion at the time of reporting and may not be the result of extensive investigation, their presence represent a richness in crash data.

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				
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

Figure 9 - Contributory factors collected by police in UK crash record format.

The Great Britain crash database collects crash severity with three different levels: slight injury, serious injury, and fatal. It is considered fatal a crash where at least one person is killed whereas other casualties - if any - may have serious or slightly injuries. Crash severity is classified according to the injury severity of the most seriously injured person involved in the crash. It is considered killed in the crash a casualty who sustained injuries which caused death less than 30 days after the crash, it is considered serious injury an injury for which a person is detained in hospital as an "in-patient", or any of the following injuries: fractures, concussion, internal injuries, burns, severe cuts, severe general shock requiring medical treatment and injuries causing death 30 or more days after the crash, and lastly, it is considered a slight injury an injury of a minor character such as a sprain (including neck whiplash injury), bruise or cut which are not judged to be severe, or slight shock requiring roadside attention including injuries which medical treatment are not required. Thus, crash severities were as follows: fatal (n = 1,366; 2.0% of the total crashes), serious (n = 16,359; 24.3% of the total crashes), and slight (n = 49,631; 73.7% of the total crashes). A small proportion of fatal crashes is a common feature of crash datasets, hence many researchers merge fatal crashes with severe crashes to gain better performance from the implemented models. In this research, despite the small share of fatal crashes, fatal and serious injury crashes were not merged in order to identify both contributory factors of fatal crashes as well as contributory factors of serious injury crashes.

Below, the pedestrian crash database is reported from Table 28 to Table 33. The variables were divided into crash (Parts A e B), vehicle (Parts A e B), driver and pedestrian sections. Several variable categories were aggregated and recoded to avoid extremely small occurrences of some categories, to remove redundant information and, finally, make the models easier to interpret.

As far as crash general condition, descriptive statistics show some categories with higher crash severities, such as motorway road class, speed limit equal to or greater than 50 mph, darkness site condition without lighting, fog or mist weather even though adverse weather such as in presence of high wind seem increase the occurrence of the most severe pedestrian crashes, weekend days. On rural roads, which are defined as roads within an area of population under 10,000, occurred the most serious crashes (26.9% of crashes in rural area) and the majority of fatalities (5.7% of crashes in rural area) despite the fact that the majority of casualties occurred on urban roads (88.1% of total crashes). Old as well as articulated vehicles, skidding vehicles and front pedestrian-vehicle first impact seem being potential factors affecting crash severity. Other categories with higher crash severities regarded very young drivers and male drivers and very old pedestrians and at the centre of carriageway as pedestrian location.

Table 28 – GB descriptive statistics related to crash data (Part A).

Variable	Fatal		Serious		Slight		Total	
	N	%	N	%	N	%	N	%
1st Road Class								
Motorway	47	32.2	49	33.6	50	34.2	146	0.2
A	747	3.3	5,941	26.2	16,013	70.5	22,701	33.7
B	128	1.8	1,819	25.7	5,129	72.5	7,076	10.5
C	78	1.6	1,067	22.0	3,712	76.4	4,857	7.2
Missing	366	1.1	7,483	23.0	24,727	75.9	32,576	48.4
Road Type								
Dual carriageway	296	5.2	1,653	28.9	3,763	65.9	5,712	8.5
Single carriageway	990	1.8	13,285	24.4	40,200	73.8	54,475	80.9
One way street	43	1.1	833	21.3	3,026	77.5	3,902	5.8
Roundabout	15	1.4	236	21.5	846	77.1	1,097	1.6
Slip road	12	2.4	97	19.6	387	78.0	496	0.7
Missing	10	0.6	255	15.2	1,409	84.2	1,674	2.5
2nd Road Class								
Motorway	5	17.9	9	32.1	14	50.0	28	0.0
A	97	1.8	1,284	23.6	4,051	74.6	5,432	8.1
B	46	2.3	492	24.5	1,471	73.2	2,009	3.0
C	34	1.6	486	22.6	1,631	75.8	2,151	3.2
Missing	439	1.7	6,553	24.7	19,574	73.7	26,566	39.4
na	745	2.4	7,536	24.2	22,891	73.4	31,172	46.3
Speed Limit								
20	74	0.9	1,840	21.9	6,476	77.2	8,390	12.5
30	821	1.5	13,007	23.9	40,697	74.6	54,525	81.0
40	129	5.4	829	34.7	1,429	59.9	2,387	3.5
≥50	342	16.7	681	33.3	1,020	49.9	2,043	3.0
Missing		0.0	2	18.2	9	81.8	11	0.0
Junction Detail								
T or staggered junction	366	1.7	5,472	24.8	16,240	73.6	22,078	32.8
Crossroads	108	1.9	1,411	24.6	4,208	73.5	5,727	8.5
More than 4 arms (not roundabout)	14	1.6	199	23.4	638	75.0	851	1.3
Mini-roundabout	6	1.0	128	21.5	462	77.5	596	0.9
Roundabout	34	1.8	467	24.1	1,438	74.2	1,939	2.9
Slip road	27	7.2	103	27.5	244	65.2	374	0.6
Private drive or entrance	25	1.7	325	21.9	1,135	76.4	1,485	2.2
Not at junction	745	2.4	7,536	24.2	22,891	73.4	31,172	46.3
Other junction	41	1.5	697	25.0	2,051	73.5	2,789	4.1
Missing		0.0	21	6.1	324	93.9	345	0.5
Junction Control								
Authorised person	2	0.6	60	17.9	273	81.5	335	0.5
Auto traffic signal	163	2.1	1,939	25.5	5,514	72.4	7,616	11.3
Give way/uncontrolled	451	1.7	6,669	24.8	19,792	73.5	26,912	40.0
Stop sign	3	0.9	64	19.9	254	79.1	321	0.5
Not at junction or within 20 metres	747	2.3	7,627	23.7	23,798	74.0	32,172	47.8

Table 29 – GB descriptive statistics related to crash data (Part B).

Variable	Fatal		Serious		Slight		Tot	
	N	%	N	%	N	%	N	%
Area								
Rural	457	5.7	2,149	26.9	5,392	67.4	7,998	11.9
Urban	909	1.5	14,208	23.9	44,232	74.5	59,349	88.1
Missing	-	0.0	2	22.2	7	77.8	9	0.0
Pedestrian Crossing Human Control								
School crossing patrol	2	0.4	88	17.8	403	81.7	493	0.7
None within 50 metres	1,345	2.1	15,918	24.6	47,494	73.3	64,757	96.1
Other	14	1.3	232	21.7	824	77.0	1,070	1.6
Missing	5	0.5	121	11.7	910	87.8	1,036	1.5
Pedestrian Crossing Physical Facilities								
No physical crossing facilities within 50 metres	931	2.1	10,567	24.1	32,387	73.8	43,885	65.2
Central refuge	67	2.7	702	28.1	1,725	69.2	2,494	3.7
Footbridge/subway	8	6.2	48	36.9	74	56.9	130	0.2
Pedestrian phase at traffic signal junction	125	1.8	1,785	25.4	5,108	72.8	7,018	10.4
Pelican, puffin, toucan or similar non-junction pedestrian light crossing	192	2.5	2,102	27.4	5,368	70.1	7,662	11.4
Zebra	39	0.8	1,038	20.4	4,005	78.8	5,082	7.5
Missing	4	0.4	117	10.8	964	88.8	1,085	1.6
Lighting								
Daylight	632	1.3	10,840	22.8	36,040	75.9	47,512	70.5
Darkness - lighting unknown	31	2.2	300	21.6	1,056	76.1	1,387	2.1
Darkness - lights lit	456	2.7	4,654	27.9	11,585	69.4	16,695	24.8
Darkness - lights unlit	25	4.9	151	29.3	339	65.8	515	0.8
Darkness - no lighting	222	17.8	414	33.2	611	49.0	1,247	1.9
Weather								
Fine no high winds	1,127	2.1	13,423	24.4	40,369	73.5	54,919	81.5
Fine + high winds	17	2.7	180	29.0	423	68.2	620	0.9
Fog or mist	8	5.0	45	28.3	106	66.7	159	0.2
Raining + high winds	21	3.1	208	31.1	440	65.8	669	1.0
Raining no high winds	137	2.0	1,693	25.3	4,857	72.6	6,687	9.9
Snowing	13	3.4	101	26.2	272	70.5	386	0.6
Other	17	1.4	253	21.3	916	77.2	1,186	1.8
Missing	26	1.0	456	16.7	2,248	82.3	2,730	4.1
Pavement								
Dry	921	1.8	12,158	23.8	37,997	74.4	51,076	75.8
Wet or damp	432	2.9	3,914	26.6	10,393	70.5	14,739	21.9
Snowy/Frozen	12	1.7	173	24.7	515	73.6	700	1.0
Missing	1	0.1	114	13.6	726	86.3	841	1.2
Day of Week								
Weekday	955	1.8	12,413	23.7	39,094	74.5	52,462	77.9
Weekend	411	2.8	3,946	26.5	10,537	70.7	14,894	22.1
Crash Severity	1,366	2.0	16,359	24.3	49,631	73.7	67,356	100.0

Table 30 – GB descriptive statistics related to vehicle data (Part A).

Variable	Fatal		Serious		Slight		Tot	
	N	%	N	%	N	%	N	%
Number of Vehicles								
1	1,170	1.9	15,171	24.1	46,635	74.1	62,976	93.50
2	143	3.9	958	25.9	2,603	70.3	3,704	5.50
>2	53	7.8	230	34.0	393	58.1	676	1.00
Vehicle Type								
Bicycle	8	0.6	399	28.2	1,006	71.2	1,413	2.10
PTW<500	23	0.9	614	24.9	1,833	74.2	2,470	3.67
PTW≥500	32	4.7	206	30.2	445	65.2	683	1.01
Car	906	1.7	12,789	23.9	39,724	74.4	53,419	79.31
Van	92	2.3	1,033	25.3	2,960	72.5	4,085	6.06
Bus	72	2.6	704	25.6	1,976	71.8	2,752	4.09
Truck	199	13.6	375	25.7	885	60.7	1,459	2.17
Other	27	3.4	187	23.3	587	73.3	801	1.19
Missing	7	2.6	52	19.0	215	78.5	274	0.41
Vehicle Towing and Articulation								
Articulated vehicle	97	28.9	110	32.7	129	38.4	336	0.50
No tow/articulation	1,252	1.9	15,989	24.4	48,280	73.7	65,521	97.28
Other	13	4.7	83	29.7	183	65.6	279	0.41
Missing	4	0.3	177	14.5	1,039	85.2	1,220	1.81
Vehicle Manoeuvre								
Going ahead	1,060	2.7	10,717	26.9	28,032	70.4	39,809	59.10
Turning left/right/U	101	1.1	2,127	23.6	6,770	75.2	8,998	13.36
Moving off	67	1.3	961	19.3	3,943	79.3	4,971	7.38
Overtaking	30	1.3	573	24.3	1,755	74.4	2,358	3.50
Reversing	61	1.2	964	19.1	4,033	79.7	5,058	7.51
Other	42	0.9	851	18.4	3,738	80.7	4,631	6.88
Missing	5	0.3	166	10.8	1,360	88.8	1,531	2.27
Vehicle Location								
At junction	620	1.8	8,711	24.9	25,691	73.4	35,022	52.00
Not at junction	744	2.4	7,533	24.2	22,895	73.4	31,172	46.28
Missing	2	0.2	115	9.9	1,045	89.9	1,162	1.73

Table 31 – GB descriptive statistics related to vehicle data (Part B).

Variable	Fatal		Serious		Slight		Tot	
	N	%	N	%	N	%	N	%
Vehicle Skidding and Overturning								
No	1,222	1.9	15,508	24.3	47,089	73.8	63,819	94.75
Yes	141	7.6	654	35.4	1,054	57.0	1,849	2.75
Missing	3	0.2	197	11.7	1,488	88.2	1,688	2.51
Vehicle 1st Point of Impact								
Back	63	1.2	1,031	19.4	4,230	79.5	5,324	7.90
Front	1,041	2.7	9,932	26.1	27,023	71.1	37,996	56.41
Nearside/Offside	219	1.1	4,577	23.4	14,755	75.5	19,551	29.03
No impact	35	1.1	631	20.4	2,431	78.5	3,097	4.60
Missing	8	0.6	188	13.5	1,192	85.9	1,388	2.06
Vehicle Engine (CC)								
<1000	100	2.1	1,271	27.0	3,336	70.9	4,707	6.99
1000-1500	236	1.8	3,426	25.7	9,692	72.6	13,354	19.83
1500-2000	417	1.9	5,456	25.3	15,675	72.7	21,548	31.99
2000-3000	155	2.5	1,594	25.8	4,435	71.7	6,184	9.18
>3000	233	6.9	932	27.7	2,204	65.4	3,369	5.00
Missing	225	1.2	3,680	20.2	14,289	78.5	18,194	27.01
Vehicle Propulsion Code								
Heavy oil	650	2.9	5,869	26.2	15,886	70.9	22,405	33.26
Hybrid electric	14	1.0	258	17.7	1,184	81.3	1,456	2.16
Petrol	479	1.9	6,537	25.9	18,244	72.2	25,260	37.50
Other	2	1.0	60	29.0	145	70.0	207	0.31
Missing	221	1.2	3,635	20.2	14,172	78.6	18,028	26.77
Vehicle Age								
≤15	1,002	2.3	11,292	25.6	31,869	72.2	44,163	65.57
>15	79	2.6	853	28.3	2,079	69.0	3,011	4.47
Missing	285	1.4	4,214	20.9	15,683	77.7	20,182	29.96

Table 32 – GB descriptive statistics related to driver data.

Variable	Fatal		Serious		Slight		Tot	
	N	%	N	%	N	%	N	%
Driver Journey Purpose								
Commuting to/from work	147	2.5	1,759	30.1	3,944	67.4	5,850	8.69
Journey as part of work	399	3.4	3,107	26.3	8,299	70.3	11,805	17.53
To/from school	7	0.4	317	19.8	1,277	79.8	1,601	2.38
Other	108	2.6	1,387	33.4	2,653	64.0	4,148	6.16
Missing	705	1.6	9,789	22.3	33,458	76.1	43,952	65.25
Driver Gender								
F	217	1.3	3,917	24.2	12,050	74.5	16,184	24.03
M	1,079	2.7	10,503	26.2	28,529	71.1	40,111	59.55
Missing	70	0.6	1,939	17.5	9,052	81.8	11,061	16.42
Driver Age								
≤ 24	194	2.8	2,062	29.3	4,776	67.9	7,032	10.44
25-34	284	2.3	3,215	26.3	8,718	71.4	12,217	18.14
35-44	230	2.2	2,627	25.2	7,550	72.5	10,407	15.45
45-54	242	2.4	2,548	25.5	7,191	72.0	9,981	14.82
55-64	187	2.7	1,800	26.2	4,887	71.1	6,874	10.21
65-74	95	2.5	987	26.0	2,713	71.5	3,795	5.63
≥ 75	60	2.3	740	28.6	1,785	69.1	2,585	3.84
Missing	74	0.5	2,380	16.5	12,011	83.0	14,465	21.48
Driver IMD Decile								
Less deprived	441	2.7	4,432	27.0	11,570	70.4	16,443	24.41
More deprived	542	2.2	6,652	26.4	17,959	71.4	25,153	37.34
Missing	383	1.5	5,275	20.5	20,102	78.0	25,760	38.24
Driver Home Area								
Rural	126	3.6	995	28.6	2,357	67.8	3,478	5.16
Small town	108	3.4	922	29.4	2,109	67.2	3,139	4.66
Urban	899	2.3	10,462	26.3	28,415	71.4	39,776	59.05
Missing	233	1.1	3,980	19.0	16,750	79.9	20,963	31.12

Table 33 – GB descriptive statistics related to pedestrian data.

Variable	Fatal		Serious		Slight		Tot	
	N	%	N	%	N	%	N	%
Number of pedestrian Involved								
1	1,280	2.0	15,691	24.0	48,301	74.0	65,272	96.91
2	66	3.6	572	30.8	1,220	65.7	1,858	2.76
>2	20	8.8	96	42.5	110	48.7	226	0.34
Pedestrian Gender								
F	458	1.6	6,864	23.2	22,216	75.2	29,538	43.85
M	908	2.4	9,494	25.1	27,406	72.5	37,808	56.13
Missing	-	0.0	1	10.0	9	90.0	10	0.01
Pedestrian Age								
0-14	67	0.4	3,442	22.9	11,516	76.6	15,025	22.31
15-24	148	1.3	2,505	21.5	9,002	77.2	11,655	17.30
25-34	160	1.6	2,049	20.9	7,593	77.5	9,802	14.55
35-44	155	2.1	1,578	21.1	5,732	76.8	7,465	11.08
45-54	153	2.1	1,694	23.7	5,306	74.2	7,153	10.62
55-64	151	2.7	1,551	27.6	3,919	69.7	5,621	8.35
65-74	152	3.4	1,494	33.4	2,826	63.2	4,472	6.64
≥75	379	7.5	1,897	37.3	2,803	55.2	5,079	7.54
Missing	1	0.1	149	13.7	934	86.2	1,084	1.61
Pedestrian Location								
Crossing elsewhere within 50 m of pedestrian crossing	118	2.1	1,511	27.5	3,866	70.4	5,495	8.16
Crossing on pedestrian crossing facility	182	1.7	2,518	24.1	7,727	74.1	10,427	15.48
In carriageway, crossing elsewhere	516	1.8	7,500	25.9	20,968	72.3	28,984	43.03
In carriageway, not crossing	220	3.2	1,449	20.9	5,272	76.0	6,941	10.30
In centre of carriageway	90	3.1	769	26.6	2,034	70.3	2,893	4.30
On footway or verge	125	1.8	1,398	20.7	5,238	77.5	6,761	10.04
Missing	115	2.0	1,214	20.7	4,526	77.3	5,855	8.69
Pedestrian Movement								
Crossing from driver's nearside	440	2.0	5,742	25.5	16,367	72.6	22,549	33.48
Crossing from driver's offside	315	2.3	3,717	26.8	9,863	71.0	13,895	20.63
Crossing from nearside - masked by parked or stationary vehicle	19	0.4	1,199	26.3	3,344	73.3	4,562	6.77
Crossing from offside - masked by parked or stationary vehicle	30	1.0	839	27.1	2,222	71.9	3,091	4.59
In carriageway, stationary - not crossing(standing or playing)	69	2.1	598	18.5	2,565	79.4	3,232	4.80
In carriageway, stationary - not crossing- masked by parked or stationary vehicle	8	1.5	112	21.6	399	76.9	519	0.77
Walking along in carriageway, back to traffic	64	4.3	329	21.9	1,109	73.8	1,502	2.23
Walking along in carriageway, facing traffic	40	4.2	200	21.0	711	74.8	951	1.41
Missing	381	2.2	3,623	21.2	13,051	76.5	17,055	25.32
Pedestrian IMD Decile								
Less deprived	412	2.4	4,207	24.8	12,311	72.7	16,930	25.14
More deprived	541	1.6	7,999	24.1	24,713	74.3	33,253	49.37
Missing	413	2.4	4,153	24.2	12,607	73.4	17,173	25.50

4.2 Swedish data

The Swedish Transport Agency collects and provides statistics on road traffic crashes storing the information in the Swedish crashes and injury database STRADA (Swedish Traffic Accident Data Acquisition) which is a national information system also based on Geographic Information Systems (GIS). STRADA was first created and implemented in cooperation with the Swedish Police, the Federation of Swedish County Councils, the National Board of Health and Welfare, the Swedish Association of Local Authorities, the Swedish Institute for Transport and Communications Analysis (SIKA) and Statistics Sweden (SCB). However, the Swedish Transport Agency is actually the authority responsible for STRADA.

Since 2016, the statistics are based on data reported by two sources, the police and emergency hospitals. Police collects all traffic crashes involving personal injury since 2003 whereas all Swedish emergency hospitals information on people who have sought care for road crash injury has been nationwide since 2016. Both authorities run the coordinated national registration of crashes and road injuries collecting data about the injured persons and the crash, at the crash site and in the emergency room, respectively, using different questionnaires. Crash data can be retrieved on STRADA website. However, to gain access to the database, an authorization by the Swedish Transport Agency is required. After setting search conditions, the website presents the information for each crash separated into police report and/or report from health care.

Police report provides time, place, crash type, crash description, weather conditions, road surface conditions, type of environment, light conditions, location type, attributes like road number, speed regulation, traffic elements and involved road users. The traffic elements refer to the vehicles (or cyclists, pedestrians) that are involved in the crash. The police register the role that each element played in the crash. A traffic element can have one or more persons linked to it. The police classify each person involved in a traffic accident as either slightly injured, severely injured, or killed. A driver can also be classified as uninjured whereas uninjured passengers are generally not included in the police report. In the database, there is also included a police sketch describing the crash. The sketch can look like the figures below.

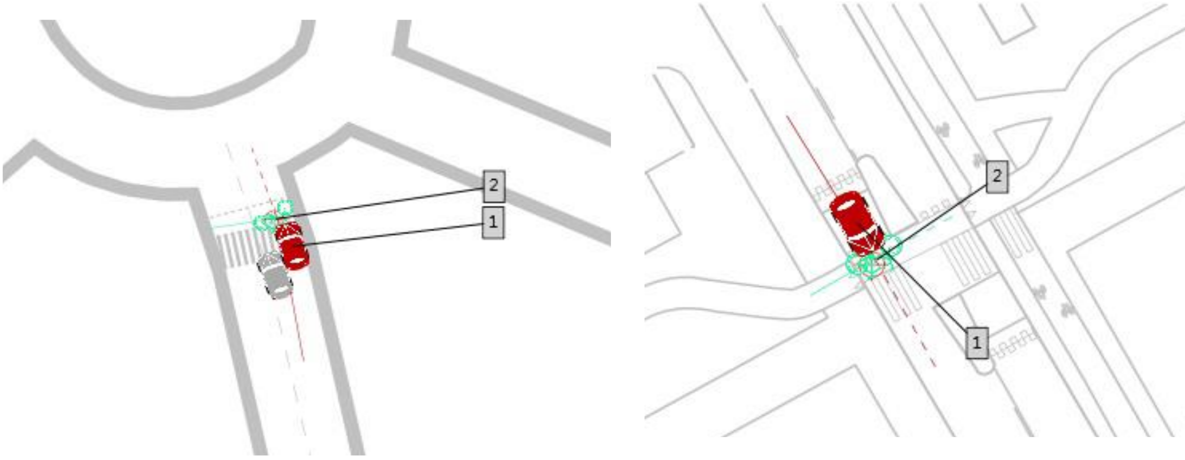


Figure 10 – Examples of police crash sketches in STRADA.

Healthcare reports (provided by the hospitals) contain supplementary data such as diagnosis classified according to Abbreviated Injury Scale (AIS), Maximum Abbreviated Injury Scale (MAIS), Injury Severity Score (ISS), International Classification of Diseases (ICD 10) and Reaction Level Scale (RLS). It is considered a slightly injured person any person slightly injured in road traffic crashes reported by the police. A road fatality is defined as any person killed in a traffic crash, or who dies within 30 days as a result of injuries sustained in the crash (this is also the common definition of fatal crashes adopted by other European countries).

For serious injuries, two definitions are used. For generating official statistics, road traffic crashes with fatal and severe personal injuries reported by the police are used. For preventive road safety work, the definition of serious injury is based on health loss following a traffic injury. If the individual does not recover after a certain amount of time, they are defined as seriously injured. Nevertheless, Sweden does not use the score of three or more on the Maximum Abbreviated Injury Scale (MAIS3+) as a formal measure of a seriously injured person. MAIS3+ is, however, used to calculate the number of persons seriously injured and is therefore an important part of the Swedish efforts to increase the level of road safety (IRTAD, 2020).

In addition, the hospital reports are centred on the person seeking medical care, as opposed to the police reports that are centred on the crash. Consequently, the health system catches many of the unprotected road users that the Police do not become aware of, such as pedestrians, cyclists and motorcyclists. Healthcare report also provides a sketch, looking like Figure 11, showing the injuries and the severity of the user according to AIS and ISS. The colour grading at the top gives the severity. The level of severity increases from the left to the right.

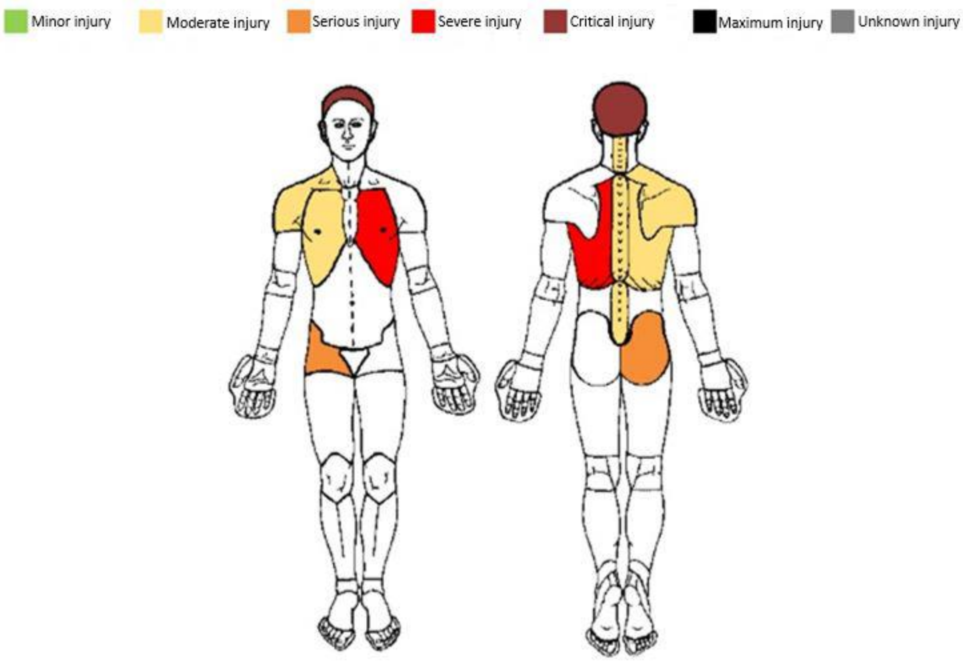


Figure 11 – Example of hospital injury sketch in STRADA.

The data are entered by the local police and hospitals into STRADA, where a match immediately takes place with a high level of accuracy for each crash.

It is noteworthy to observe that police produce one report per crash, whereas the hospitals produce one report per person involved.

Further information about crash reporting system is available at https://www.transportstyrelsen.se/sv/vagtrafik/statistik/olycksstatistik/. The STRADA report form filled in by the police actually is provided below (from Figure 12 to Figure 14).

Once in STRADA, the data which can be retrieved are organized in five different files over the first file which contains just summary crash statistics regarding when the crash data were retrieved, the number of crashes collected in the years chosen as study period and the relative crash severity. Thus, the first file containing data is named “Crashes” and specifically contains general crash info. For each record, it is reported if (and how many) police and healthcare reports were registered, the year, month, and day of the crash, the county, the municipality, the area, the site where the crash took place, the type of crash, a brief description of the course of events, the crash severity, road and environmental information such as the road conditions, the lighting condition, the road type and the road holder, the speed limit, the vehicles involved in the crash (indicated as traffic elements).

Another file named “People” contains info related to the people involved in the crash. The spreadsheet contains a large number of columns with values from either the police or from the healthcare system. Hence, chunks of information are repeated. However, variables describing the person involved were further provided (the age and gender, the eventual alcohol or drug use, the role in the crash – pedestrian, driver, passenger – and different columns referring to the severity (ISS, MAIS, and a weighted degree of damage). There is also a set of variables describing if the road user was in conflict with another user. However, the latter were used to double-check data after the pre-processing phase. “Crashes” and “People” files were the results of police and hospital reports which main information can be consulted as well. Finally the last file contains detailed information related to the body damage reported by the hospitalised people.

In cases where a value can be obtained either from the police or from the health service, the police's tasks usually take precedence. This applies, for example, to information about the crash site or crash type. If the police and the health service have stated different information about the degree of injury of the people involved, the health care's information is usually given priority.

Polisen

Polismyndigheten

INFORMATIONSUNDERLAG
Vägtrafikolycka

Sida 1 (1)

Datum

Kopia ska lämnas för registrering i STRADA. Glöm ej diarienummer

Telefon dagtid | 52 Polisens diarienummer

A Nr

Polisområdeskod	54 Kommun	55 Tidpunkt för olyckan	Datum	Klockslag	Veckodag	Väghållarkod

56 Olycksplats (ange gatu-/vägnamn/vägnr, ev husnr samt avstånd till närmaste korsning mellan allmänna vägar)

Namn på stadsdel/kommundel/ort el dyl

57 Skiss, på vilken anges gatu- och vägnamn, vägbredd, åtföljd av bokstav A resp B enl. avsnitt B nedan. Vid inritat fordon anges fordonsslag (pb, lb, etc) ett trafikelement - (vägtrafikant-) nummer 1, 2, 3 osv, vilket nr skall vara identiskt med det nr vederbörande vägtrafikant åsatts i trafikmålsanteckningar (RPS 411.20)

Norrpil

58 Kortfattad beskrivning av händelseförloppet, siktförhållanden m.m.

B Väg och trafik

	Väg A	Väg B
59 Vägnummer		
60 Högsta tillåtna hastighet		
61 Vägtyp		
Motorväg 1		
Motortrafikled 2		
Annan allm väg 3		
Gata 4		
Enskild väg 5		
Övr väg, torg, etc 6		

Trafikanvisningar *)		Väg A	Väg B
Huvudled	1		
Ej huvudled	2		
65 Trafikreglering *)			
Förb mot v-sväng	1		
Stopplikt	2		
Väjningsplikt	3		
64 Trafiksignal *)			
I funktion	1		
Ur funktion	2		
Gult blinkande	3		
Saknas	4		

C Väderlek, väglag, belysning

63 Väderleksförhållanden	
Uppehållsväder	1
Dis/dimma	2
Regn	3
Snöblandat regn	4
Snöfall	5
66 Väglag	
Vägbanan torr	1
Vägbanan våt/fuktig	2
Tjock is/packad snö	3
Tunn is (vägb synlig)	4
Lös snö/snömodd	5

67 Trafikmiljö			
Tätbebyggt område	1		
Ej tätbebyggt område	2		
68 Ljusförhållanden			
Dagsljus	1		
Mörker	2		
Gryning/skymning	3		
Om 68:2 eller 3 förkryssats			
69 Gatu-/vägbelysning		Väg A	Väg B
Tänd	1		
Släckt	2		
Saknas	3		

D Trafikelement / E Inblandade personer

70 Nr	Trafikelement (t.ex. pb, lätt/tung lb, lätt/ tung mc, cykel, gående enl. 1 Kap. 4 § TaF, vilt/djur)	Registreringsnr. (anges för motor och släpfordon). För utländskt fordon, nationalitet	Totalt antal pers i fordonet	Övningskörning **) Trafik skola	Övningskörning **) Privat	71 Födelsetid Obligatoriskt för förare och instruktör samt dödade och skadade personer åååå mm dd	72 Trafikant Förare el. elev som kör. Ange F/E	Passagerare/ instruktör Fram	Passagerare/ instruktör Bak	Passagerare/ instruktör Okänt eller övrigt	Död	Svårt skadad	Lindrigt skadad	Bälte Ange J/N	74 Misstänkt påverkad av alkohol/ annat ämne (förare) Ange J/N

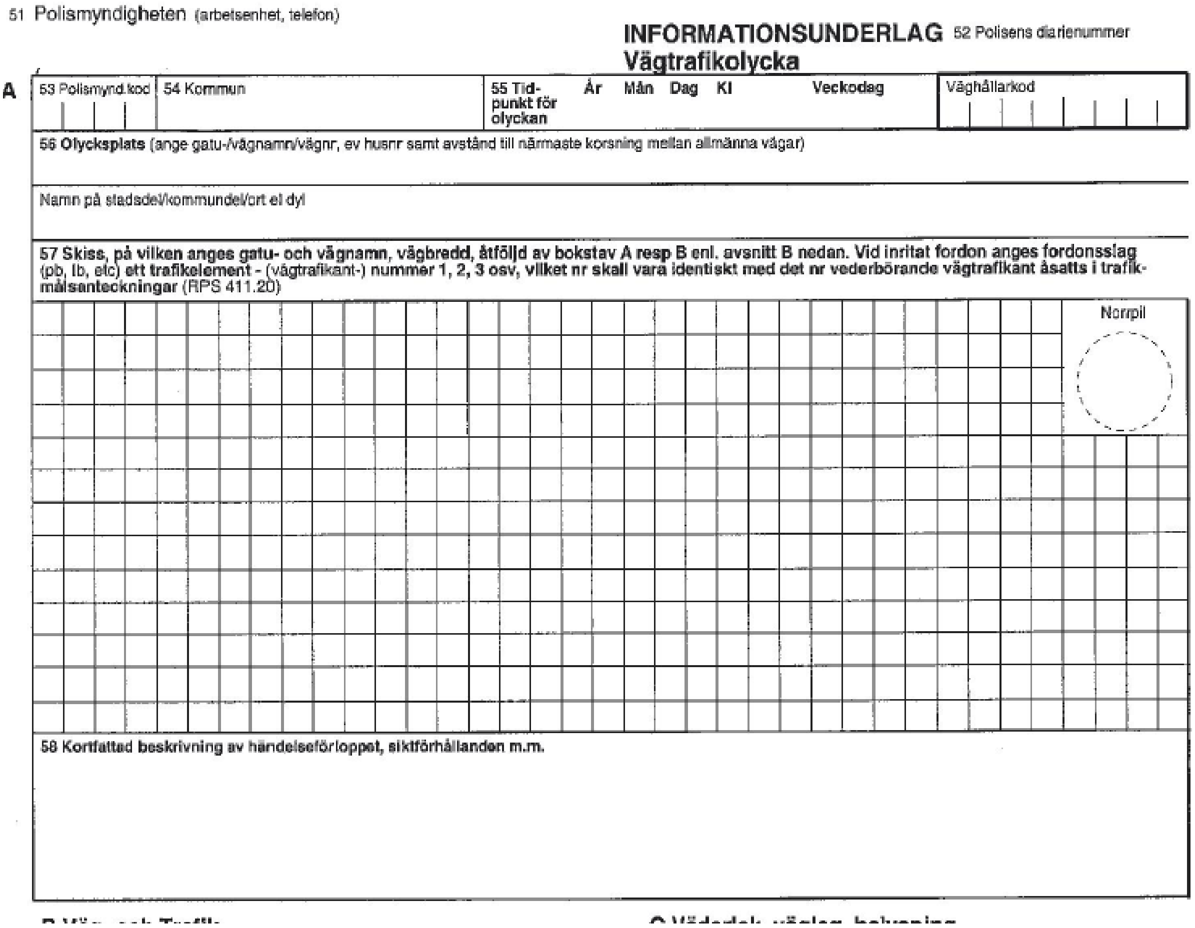
Fordon skyltat för transport av farligt gods inblandat. Ange elementnr:

Datum	Ort	Uppgiftslämnare

50 Statistiska uppgifter registrerat i STRADA olycksdatabas ☐ Transportstyrelsen	Datum	Sign

*) Kontrolleras **) Med övningskörning avses enbart de fall då eleven framfört fordonet, alltså ej då instruktören kört.

Figure 12 – Swedish police crash report form (part A).



51 Polismyndigheten (arbetsenhet, telefon)

INFORMATIONSUNDERLAG 52 Polisens diarienummer
Vägtrafikolycka

A 53 Polismynd.kod | 54 Kommun | 55 Tid-punkt för olyckan | År | Mån | Dag | Kl | Veckodag | Väghållarkod

56 Olycksplats (ange gatu-/vägnamn/vägnr, ev husnr samt avstånd till närmaste korsning mellan allmänna vägar)

Namn på stadsdel/kommundel/ort el dyl

57 Skiss, på vilken anges gatu- och vägnamn, vägbredd, åtföljd av bokstav A resp B enl. avsnitt B nedan. Vid inritat fordon anges fordonsslag (pb, lb, etc) ett trafikelement - (vägtrafikant-) nummer 1, 2, 3 osv, vilket nr skall vara identiskt med det nr vederbörande vägtrafikant åsatts i trafikmålsanteckningar (RPS 411.20)

Norrpil

58 Kortfattad beskrivning av händelseförloppet, siktförhållanden m.m.

Informationsunderlag, vägtrafikolycka	Information basis, road traffic accident
51 Polismyndigheten (arbetsenhet, telefon)	51 Police authority (work unit, telephone)
52 Polisens diarienummer	52 Registration number at Police *(unique for each accident)*
53 Poiismynd kod	53 Code for Police authority
54 Kommun	54 Municipality *(there are 290 municipalities in Sweden. The include cities and towns)*
55 Tidpunkt för olyckan År Mån Dag Kl Veckodag	55 Time of the accident Year Month Day Time Weekday
Väghållarkod	A code for the authority or city responsible for the road or street
56 Olycksplats (ange gatu-/vägnamn/vägnr, ev husnr samt avstånd till närmaste kortsning mellan allmänna vägar)	56 Accident location (write street name/road name/road number, street number (if one) and distance to nearest intersection between two public roads)
Namn på stadsdel/kommundel/ort el dyl	Name of city area/municipality area/village or similar
57 Skiss, på vilken anges gatu- och vägnamn, åtföljd av bokstav A resp B enl. avsnitt B nedan. Vid inritat fordon anges fordonsslag (pb, lb, etc) ett trafikelement- (vägtrafikant-)nummer 1,2,3 osv, vilket nr skall vara identiskt med det nr vederbörande vägtrafikant åsatts I trafikmålsanteckningarna (RPS 411.20)	57 Drawing, on which shall be given street and road name, followed by letters A or B respectively, according to section B below. For each drawn vehicle shall be written type of vehicle (truck, pass car etc), a number – 1,2,3 etc – for each vehicle or road user and this number shall be identical to the number given for the road user in the traffic incident notes (RPS 411.20). *(traffic incident notes refer to additional notes on the accidents that may serve as part of the basis for prosecution, if that occurs)*
Norrpil	Arrow indicating north
56 Kortfattad beskrivning av händelseförloppet, siktförhållanden m m	56 Short description of accident events, sight conditions etc.

Figure 13 – Swedish police crash report form (part B).

B Väg- och Trafik | **C Väderlek, väglag, belysning**

	Väg A	Väg B	62 Trafikanvisningar*)		Väg A	Väg B	65 Väderleksförhållanden			67 Trafikmiljö		
			Huvudled	1			Uppehållsväder	1		Tättbebyggt område	1	
59 Vägnummer			Ej huvudled	2			Dis/dimma	2		Ej tättbebyggt område	2	
60 Högsta tillåtna hastighet			63 Trafikreglering*)				Regn	3		68 Ljusförhållanden		
			Förb mot v-sväng	1			Snöblandat regn	4		Dagsljus	1	
61 Vägtyp			Stopplikt	2			Snöfall	5		Mörker	2	
Motorväg 1			Väjningsplikt	3			66 Väglag			Gryning/skymning	3	
Motortrafikled 2			64 Trafiksignal*)				Vägbanan torr	1		Om 68:2 eller 3 förkryssats		
Annan allm väg 3			I funktion	1			Vägbanan våt/fuktig	2		69 Gatu-/vägbelysning	Väg A	Väg B
Gata 4			Ur funktion	2			Tjock is/packad snö	3		Tänd 1		
Enskild väg 5			Gult blinkande	3			Tunn is (vägb synlig)	4		Släckt 2		
Övr väg, torg etc 6			Saknas	4			Lös snö/snömodd	5		Saknas 3		

D Trafikelement | **E Inblandade personer**

B Road and Traffic							C Weather, road conditions, lighting					
	Road A	Road B	62 Priority rules *)		Road A	Road B	65 Weather conditions			67 Traffic environment		
			Priority road	1			No precipitation	1		Built-up area	1	
59 Road number			Non-priority road	2			Mist/fog	2		Not built-up are	2	
60 Maximum allowed speed			63 Traffic control *)				Rain	3		68 Light conditions		
			Left-turn prohibited	1			Snowy rain	4		Daylight	1	
61 Road type			Mandatory stop	2			Snowfall	5		Darkness	2	
Motorway 1			Mandatory yield	3			66 Road surface conditions			Dawn/dusk	3	
Expressway 2			64 Traffic Signal *)				Dry road surface	1		If 68:2 or 3 has been ticked		
Other public road 3			In function	1			Wet/moist road surface	2		69 Street/road lighting	Road A	Road B
Street 4			Not in function	2			Thick ice/packed snow	3		Lit 1		
Private road 5			Blinking yellow	3			Thin ice (roadway visible)	4		Not lit 2		
Other road, square etc 6			None	4			Loose snow/sleet	5		Missing 3		

D Trafikelement | **E Inblandade personer**

70 Trafikelement						71 Personnummer	72 Trafikant				73 Personskada			74
Nr	Trafikelement (t.ex. pb, lätt/tung lb, lätt/tung mc, cykel, gående enl. 1 Kap. 4 § TrF, vilt/djur)	Registreringsnr. (anges för motor- och släpfordon). För utländskt fordon, nationalitet	Totalt antal pers i fordonet	Övningskörning**) Trafikskola	Privat	Obligatoriskt för förare och instruktör samt dödade och skadade personer	Förare el. elev som kör. Ange F/E	Passagerare/instruktör Fram	Bak	Okänt eller övrigt	Död	Svårt skadad	Lindrigt skadad	Misstänkt påverkad av alkohol/annat ämne (förare). Ange J/N

Fordon skyltat för transport av farligt gods inblandat. Ange elementnr: | FU inledd, datum och tid | av (titel, för- och efternamn)

Ort och datum	75 Undersökn.ledarens beslut		Beslutsdatum
	FU inleds ej	FU nedlagd	
Uppgiftslämnare	Ej spaningsresultat	Misstänkt ej fyllt 15 år	Undersökningsledarens namnteckning/sign
	Brott kan ej styrkas	Misstänkt avliden	
50 Statistiska uppgifter till Vägverket | Datum och sign	Misstänkt oskyldig	Rapporteftergift	
	Gärningen ej brott	Spaningsuppslag saknas	

D Traffic element | **E Involved persons**

70 Traffic element						71 Persons Identity code	72 Road User				73 Person Injury			74
No	Traffic element (eg passenger car,bus, light /heavy truck, MC, cyclist, pedestrian acc. As defined in traffic regulation	Registration number (noted for motor vehicle and trailer vehicle). Nationality if foreign veh.	Total number of persons in vehicle	Driver training **) Driving school	Private	Mandatory for driver and instructor and for killed or injured persons	Driver or driving Trainee. Note 'D' or 'T'	Passenger / Instructor Front	Back	Unknown or other	Killed	Seriously injured	Slightly injured	Suspected influence of alcohol or other substance. Write Y or N

Vehicle signed as transport of hazardous goods involved. Write element number | Initial investigation begun, date and time | By (position/title, first name and family name)

Place and date	75 Investigators decision		Date of decision
	Investigation will not be initiated	Investigation cancelled	
Information provided by	No result of surveillance	Suspect has not reached age of 15	Investigators signature
	Violation cannot be proven	Suspect deceased	
50 Statistical information to | Date & signature	Suspect innocent	Report not necessary	
Road Administration	Action was no violation	No basis for investigation	

Figure 14 – Swedish police crash report form (part C).

In this research, data related to the five-year period from 2015 to 2019 were used. Crash data files were merged into a single one by using the crash-index which is unique for each observation. The Swedish database also contains pedestrian crashes without any kind of vehicle involved. However, in uniformity with the British and the Italian crash databases and to carry out analyses with homogenous data, only pedestrian-vehicle crashes were considered. The dataset made up only by pedestrian-vehicle crashes contained 9,697 observations. In the Swedish database crash severity can assume six different categories of severity: 1) Fatal crashes, 2) Serious crashes (ISS 9+), 3) Moderate crashes (ISS 4 - 8), 4) Mild crashes (ISS 1 - 3), 5) Uncertain or unknown severity, 6) No personal injury crashes. Moreover, the Swedish database also provides property damage only (PDO) crashes. Nevertheless, the aims of this research are to understand which factors may contribute to pedestrian crash severity and compare methods according their performances. Thus, PDO crashes, which represents less than 3% of all crashes, were removed from the final dataset. The dependent variable "crash severity" was rearranged as a three level variable based on both police and hospital reports and regards only pedestrian in collision with a vehicle. In this research, the six classes collapsed in three categories of severity as follows: 1) Fatal crashes, 2) Serious injury crashes, and 3) Slight injury crashes which include moderate crashes, mild crashes, and uncertain or unknown severity.

Table 34 – Crash severity for Swedish crashes.

Crash Severity	**N**	**%**
Fatal	212	2.19
Serious	426	4.39
Slight	8,788	90.63
PDO	271	2.79
Total	**9,697**	**100.00**

Finally, the Swedish dataset contained 9,426 pedestrian-vehicle crashes with crash severity as follows: Fatal (N=212, 2.25%), serious injury (N=426, 4.52%) and slight injury (N=8,788, 93.23%). In the database, the variable lighting had numerous missing value. Thus, to not to exclude the information from the set of variables used in the analysis a new variable for lighting was created by creating an R script. Using the package "SUNCALC", the script needs as input variables the latitude and longitude of the place at which the crash occurred (which is known as in the crash database a variable indicating the county is reported). Then the tool provides the exact hour of dawn and dusk in each county. The new lighting variable was created by comparing the time of the crash and the time of dawn and dusk.

Although a considerable share of pedestrian crashes occurred in urban area, the most serious crashes were recorded in rural area (74 fatal crashes and 53 serious pedestrian crashes out of 1,023 occurred in rural area whereas in urban area they were 472 out of 7,543, 15% of fatal and serious crashes in rural area out of 5% in urban area). Furthermore, most crashes occurred on roads with speed limit up to 50 km/h. However, it was in presence of speed limits greater than or equal to 60 km/h. Descriptive statistics show other categories with higher crash severities, such as motorways, interchanges, drivers or pedestrians which have drinking, and older pedestrian age (≥75).

Table 35 – SW descriptive statistics related to crash data.

Variable	Fatal		Serious		Slight		Total	
	N	%	N	%	N	%	N	%
Area								
Rural	74	7.23	53	5.18	896	87.59	1,023	10.85
Urban	125	1.66	347	4.60	7,071	93.74	7,543	80.02
Missing	13	1.51	26	3.02	821	95.47	860	9.12
Road Type								
Motorway	20	24.39	5	6.10	57	69.51	82	0.87
Rural Individual	3	4.41	3	4.41	62	91.18	68	0.72
Rural Municipal	2	0.77	12	4.62	246	94.62	260	2.76
Rural National	45	12.82	24	6.84	282	80.34	351	3.72
Rural Other	10	3.57	9	3.21	261	93.21	280	2.97
Urban Individual	10	3.37	19	6.40	268	90.24	297	3.15
Urban Municipal	80	1.42	250	4.45	5,293	94.13	5,623	59.65
Urban National	17	4.08	31	7.43	369	88.49	417	4.42
Urban Other	12	1.01	47	3.96	1,129	95.03	1,188	12.60
Missing	13	1.51	26	3.02	821	95.47	860	9.12
Speed Limit								
≤30	16	0.80	70	3.52	1,902	95.67	1,988	21.09
40	25	1.32	109	5.74	1,765	92.94	1,899	20.15
50	61	1.97	141	4.56	2,889	93.46	3,091	32.79
≥60	77	10.49	46	6.27	611	83.24	734	7.79
Missing	33	1.93	60	3.50	1,621	94.57	1,714	18.18
Day of the week								
Weekday	169	2.26	343	4.58	6,979	93.17	7,491	79.47
Weekend	43	2.22	83	4.29	1,809	93.49	1,935	20.53
Lighting								
Daylight	121	1.96	287	4.64	5,774	93.40	6,182	65.58
Dawn dusk	9	1.82	10	2.02	475	96.15	494	5.24
Darkness	78	3.08	121	4.77	2,336	92.15	2,535	26.89
Missing	4	1.86	8	3.72	203	94.42	215	2.28
Pavement								
Dry	184	2.17	350	4.12	7,964	93.72	8,498	90.15
Wet	-	-	5	4.85	98	95.15	103	1.09
Snowy/icy	2	3.08	5	7.69	60	92.31	65	0.69
Slippery	-	-	17	7.17	218	91.98	237	2.51
Unevenness	1	1.30	8	10.39	68	88.31	77	0.82
Missing	25	5.61	41	9.19	380	85.20	446	4.73
Crash Location								
At intersection	38	1.67	105	4.61	2,134	93.72	2,277	24.16
Not at intersection	174	2.44	321	4.50	6,635	93.06	7,130	75.64
Missing	-	-	-	-	19	100.00	19	0.20
Place Type								
Road section	170	3.16	255	4.74	4,958	92.10	5,383	57.11
Intersection	34	1.66	95	4.63	1,924	93.72	2,053	21.78
Roundabout	2	0.93	9	4.19	204	94.88	215	2.28
Interchange	2	22.22	1	11.11	6	66.67	9	0.10
Pedestrian/bicycle path	4	0.29	50	3.62	1,326	96.09	1,380	14.64
Separate parking space	-	-	11	4.56	230	95.44	241	2.56
Other	-	-	-	-	19	100.00	19	0.20
Missing	-	-	5	3.97	121	96.03	126	1.34

Table 36 – SW descriptive statistics related to vehicle and users' data.

Variable	Fatal		Serious		Slight		Total	
	N	%	N	%	N	%	N	%
Vehicle Type								
Bike	1	0.07	64	4.65	1,310	95.27	1,375	14.59
Car	126	1.99	264	4.16	5,953	93.85	6,343	67.29
PTW	2	0.59	15	4.40	324	95.01	341	3.62
Truck	63	5.53	63	5.53	1,014	88.95	1,140	12.09
Other	20	8.81	20	8.81	187	82.38	227	2.41
Vehicle Trailers								
0	194	2.68	322	4.44	6,735	92.88	7,251	76.93
1	13	12.15	5	4.67	89	83.18	107	1.14
Missing	-	-	10	5.46	173	94.54	183	100.00
n.a.	5	0.27	89	4.72	1,791	95.01	1,885	20.00
Driver Gender								
Female	34	2.48	89	6.50	1,246	91.02	1,369	14.52
Male	153	4.16	208	5.65	3,318	90.19	3,679	39.03
Missing	25	0.57	129	2.95	4,224	96.48	4,378	46.45
Driver Age								
0-24	25	3.47	33	4.58	663	91.96	721	7.65
25-34	42	5.43	42	5.43	690	89.15	774	8.21
35-44	25	3.47	49	6.80	647	89.74	721	7.65
45-54	32	3.74	58	6.78	765	89.47	855	9.07
55-64	36	4.36	44	5.33	745	90.30	825	8.75
65-74	15	2.59	38	6.55	527	90.86	580	6.15
≥75	12	2.67	32	7.13	405	90.20	449	4.76
Missing	25	0.56	130	2.89	4,346	96.56	4,501	47.75
Driver Alcohol/Drug use								
No	205	2.50	411	5.00	7,599	92.50	8,215	87.15
Yes	1	14.29	3	42.86	3	42.86	7	0.07
Missing	6	0.50	12	1.00	1,186	98.50	1,204	12.77
Number of pedestrian involved								
1.00	199	2.19	409	4.49	8,492	93.32	9,100	96.54
2.00	9	3.07	15	5.12	269	91.81	293	3.11
2+	4	12.12	2	6.06	27	81.82	33	0.35
Pedestrian Gender								
Female	86	1.79	217	4.51	4,510	93.70	4,813	51.06
Male	126	3.03	190	4.57	3,839	92.39	4,155	44.08
Missing	-	-	19	4.15	439	95.85	458	4.86
Pedestrian Age								
0-14	17	1.48	41	3.57	1,089	94.94	1,147	12.17
15-24	17	1.06	30	1.87	1,557	97.07	1,604	17.02
25-34	15	1.23	28	2.30	1,173	96.46	1,216	12.90
35-44	21	2.10	35	3.50	945	94.41	1,001	10.62
45-54	20	1.86	39	3.63	1,014	94.50	1,073	11.38
55-64	26	2.65	60	6.12	894	91.22	980	10.40
65-74	33	3.91	68	8.05	744	88.05	845	8.96
≥75	63	5.92	106	9.96	895	84.12	1,064	11.29
Missing	-	-	19	3.83	477	96.17	496	5.26
Pedestrian Alcohol/Drug use								
No	98	1.11	382	4.33	8,332	94.55	8,812	93.49
Yes	7	7.00	25	25.00	68	68.00	100	1.06
Missing	107	20.82	19	3.70	388	75.49	514	5.45
Crash Severity								
Fatal	212	100.00	-	-	-	-	212	2.25
Serious	-	-	426	645.45	-	-	426	4.52
Slight	-	-	-	-	8,788	100.00	8,788	93.23

4.3 Italian data

The Italian crash database is maintained by the National Institute of Statistics (Istat) and contains all crashes occurred on a public highway (including footways, squares where traffic flow is allowed) in which at least one vehicle or a vehicle in collision with a pedestrian is involved in the crash. However, only information on injury crashes occurred on the national roads are reported. This means that PDO (property damage only) crashes are not collected by the authorities and excluded from the data as well as crashes occurred on private roads or crashes with no vehicle involvement (i.e., pedestrian only crashes). The national database is based on information collected on the scene by Highway Police, Local Police, Police, and Carabinieri (an army corp). Crash report forms used by the different police forces are different. At the same time, skills in crash reporting are quite different because both auxiliary personnel and specialized units perform similar tasks.

The database format (mod. CTT/INC, www.istat.it/it/archivio/4609) has been evolving over time and the most recent version dates back to 2019. Instructions to compile the database are provided even though they result quite synthetic and several coding errors occur. The national crash database is not linked to road and traffic database, which do not exist even though they are required by the Road Code issued in 1992.

As part of the National Highway Safety Plan implementation, some road safety monitoring centres have been set up and local crash databases have been developed. Furthermore, some municipalities are developing specific databases for urban crashes. These new databases are an improvement of the national database but there is not consistency between the different databases and they are not spread in all the states.

Since 2019, Istat has provided a new system to enhance data acquisition. Istat Gino system is indeed available to local police at https://gino.istat.it/incidenti. The local police has the access to the system where it is possible to enter online information relating to each crash detected, through the guided compilation similar to a web questionnaire. This system allows the local police and all authorised bodies to upload information by loading a .csv file, according to a record layout available at the "Documents and instructions" section at the link reported above. The adoption of this data acquisition system represents the beginning of a transition period from the previous data acquisition system to the new electronic one.

The access to the database is restricted. However, Istat organises data in two different format:

- Macrodata: information is aggregated and available and accessible on the Istat website to everyone (https://www.istat.it/it/ archive/245757).

- Microdata: all crashes data, collected into a unique database, are provided in detail. This data type is only available to particular categories of users. Microdata can be further divided into two groups: 1) research files (MFR): data created to meet scientific research needs. These are elementary data files without any direct identification elements to protect confidentiality. Access to the files may be requested only for carrying out a specific analysis by researchers belonging to recognized research bodies; 2) miIro.STA files: data for public use for specific studies. Starting from the corresponding files for research (MFR), milro.STA files are further processed for privacy protection purposes.

Thanks to the precious memorandum of understanding stipulated between Istat and the University of Naples Federico II, MFR data were used in this research and were further enriched with confidential information such as month and day of crash occurrence, pedestrian and driver behaviour and psychophysical state at the moment of the crash and the age of the people involved in the crash (in a disaggregate form). Indeed, ISTAT usually provides people involved age in age-groups.

The Istat crash report form is provided below (from Figure 15 to Figure 18).

CODICI ISTAT DELLE CIRCOSTANZE PRESUNTE DI INCIDENTE

1) CIRCOSTANZE PRESUNTE DELL'INCIDENTE per inconvenienti di circolazione

A) INCIDENTI TRA VEICOLI IN MARCIA

INCIDENTE SULL'INTERSEZIONE STRADALE (INCROCIO)

	Cod.
Procedeva regolarmente senza svoltare	01
" con guida distratta e andamento indeciso	02
" senza mantenere la distanza di sicurezza (art. 149)	03
" senza dare la precedenza al veicolo proveniente da destra (art. 145)	04
" senza rispettare lo stop (art. 145)	05
" senza rispettare il segnale di dare precedenza (art.145)	06
" contromano (art 143)	07
" senza rispettare le segnalazioni semaforiche o dell'agente (artt.41-43)	08
" senza rispettare i segnali di divieto di transito o di accesso	10
" con eccesso di velocità (art. 141)	11
" senza rispettare i limiti di velocità (art. 142)	12
" con le luci abbaglianti incrociando altri veicoli (art. 153)	13
Svoltava a destra regolarmente	14
" a destra irregolarmente	15
Svoltava a sinistra regolarmente	16
" a sinistra irregolarmente	17
Sorpassava (all'incrocio) - (art. 148)	18

INCIDENTE NON ALL'INTERSEZIONE STRADALE

	Cod.
Procedeva regolarmente	20
" con guida distratta e andamento indeciso	21
" senza mantenere la distanza di sicurezza (art. 149)	22
" con eccesso di velocità (art. 141)	23
" senza rispettare i limiti di velocità (art. 142)	24
" non in prossimità del margine destro della carreggiata (art. 143)	25
" contromano (art 143)	26
" senza rispettare i segnali di divieto di transito o di accesso	27
" con le luci abbaglianti incrociando altri veicoli (art. 153)	28
Sorpassava regolarmente	29
" irregolarmente a destra (art.148)	30
" in curva, su dosso o in condizione di insuff. visibilità (art.148)	31
" un veicolo che ne stava sorpassando un altro (art.148)	32
" senza osservare l'apposito segnale di divieto	33
Manovrava in retrocessione o conversione	34
" per immettersi nel flusso della circolazione	35
" per voltare a sinistra (passaggio privato,distributore, ecc.)	36
" regolarmente per fermarsi o sostare	37
" irregolarmente per fermarsi o sostare	38
Si affianca ad altri veicoli a due ruote irregolarmente	39

B) INVESTIMENTO DI PEDONE

VEICOLO COINVOLTO

	Cod.
Procedeva regolarmente	40
" con eccesso di velocità (art. 141)	41
" senza rispettare i limiti di velocità (art. 142)	42
" contromano (art. 143)	43
Sorpassava veicolo in marcia	44
Manovrava	45
Non rispettava le segnalazioni semaforiche o dell'agente (art. 41-43)	46
Usciva senza precauzioni da passo carrabile	47
Fuoriusciva dalla carreggiata	48
Non dava la precedenza al pedone sugli appositi attraversamenti (art. 191)	49
Sorpassava un veicolo fermatosi per consentire l'attraversamento dei pedoni	50
Urtava con il carico il pedone	51
Superava irregolarmente un tram fermo per la salita e discesa dei passeggeri	52

PEDONE INVESTITO

	Cod.
Camminava o sostava mantenendosi su marciapiede, banchina,ecc.	40
" regolarmente sul margine della carreggiata	41
" contromano (art. 190)	42
" in mezzo alla carreggiata	43
Sostava, indugiava, o giocava sulla carreggiata (art. 190)	44
Lavorava sulla carreggiata protetto da apposito segnale	45
" sulla carreggiata non protetto da apposito segnale	46
Saliva su veicolo in marcia	47
Discendeva da veicolo con prudenza	48
Discendeva da veicolo con imprudenza	49
Veniva fuori improvvisamente da dietro o davanti un veicolo in sosta o fermata	50
Attraversava la strada ad un passaggio pedonale protetto da semaforo, o da agente rispettando le segnalazioni	51
Attraversava la strada ad un passaggio pedonale protetto da semaforo, o da agente non rispettando le segnalazioni (art. 41-43)	52
Attraversava la strada ad un passaggio pedonale non protetto da semaforo o da agente	53
" la strada regolarmente, non ad un passaggio pedonale	54
" la strada irregolarmente (art. 190)	55

C) INCIDENTE A VEICOLO IN MARCIA CHE URTA VEICOLO FERMO O ALTRO OSTACOLO

VEICOLO IN MARCIA

	Cod.
Procedeva regolarmente	60
" con guida distratta e andamento indeciso	61
" senza mantenere la distanza di sicurezza (art. 149)	62
" contromano (art. 143)	63
" con eccesso di velocità (art. 141)	64
" senza rispettare i limiti di velocità (art.142)	65
" senza rispettare i segnali di divieto di transito o di accesso	66
Sorpassava un altro veicolo in marcia	67
Attraversava imprudentemente il passaggio a livello (art. 147)	68

VEICOLO FERMO O ALTRO OSTACOLO

	Cod.
Ostacolo accidentale	60
Veicolo fermo in posizione regolare	61
" in posizione irregolare (art. 158)	62
" senza che sia stato collocato il prescritto segnale (art. 162)	63
" regolarmente segnalato	64
Ostacolo fisso nella carreggiata (isole, colonnine, transenne, ecc.)	65
Treno in passaggio a livello	66

D) INCIDENTE A VEICOLO IN MARCIA SENZA URTO CON VEICOLO O OSTACOLO SULLA CARREGGIATA

VEICOLO COINVOLTO

	Cod.
Sbandamento con fuoruscita per evitare l'urto	70
" con fuoruscita per guida distratta e andamento indeciso	71
" con fuoruscita per eccesso di velocità	72
Frenata improvvisa con conseguenza ai trasportati	73
Caduta di persona da veicolo per:	
a) apertura di portiera	74
b) discesa da veicolo in moto	75
c) essersi aggrappata o sistemata inadeguatamente	76

VEICOLO, PEDONE OD OSTACOLO NON URTATI

	Cod.
Ostacolo accidentale	70
Pedone	71
Animale	72
Veicolo	73
Buche, ecc.	74
Senza ostacolo né pedone né altro veicolo	75
Ostacolo fisso	76

2) CIRCOSTANZE PRESUNTE DELL'INCIDENTE per difetti o avarie del veicolo

	Cod.
Rottura o insufficienza dei freni	80
" o guasto allo sterzo	81
Scoppio o eccessiva usura dei pneumatici	82
Mancanza o insufficienza dei fari o delle luci di posizione	83
" o insufficienza dei lampeggiatori o delle segnalazioni luminose di arresto	84
Rottura degli organi di agganciamento dei rimorchi	85
Deficienza delle attrezzature per trasporto di merci pericolose (carburante,esplosivi, gas compressi, ecc.)	86
Mancanza o insufficienza degli adattamenti prescritti per i veicoli condotti da mutilati o minorati fisici	87
Distacco di ruota	88
Mancanza o insufficienza dei dispositivi visivi dei velocipedi	89

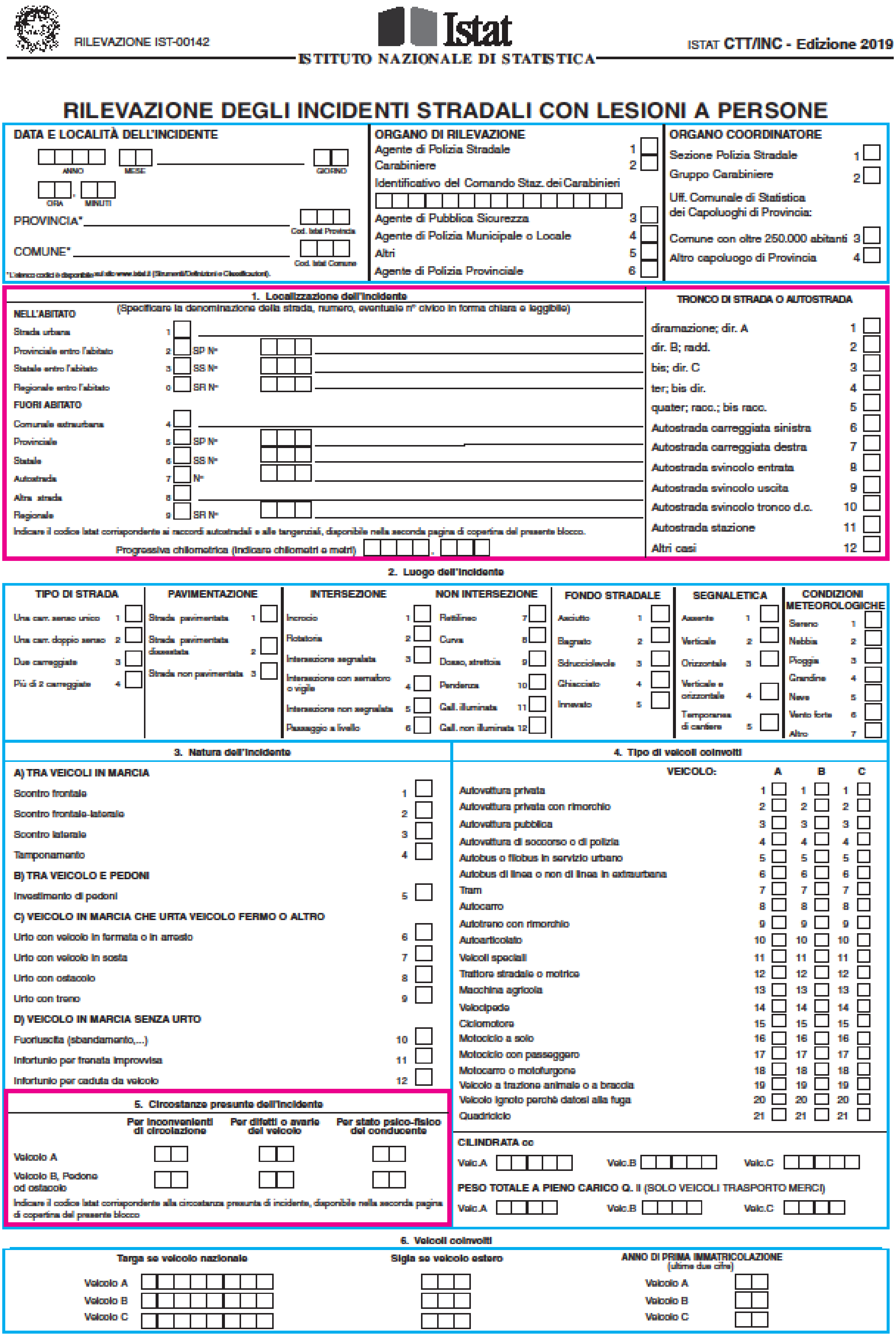
3) CONDIZIONI PRESUNTE DELL'INCIDENTE per stato psico-fisico

	Cod.
Anormale per ebbrezza da alcool (art. 186)	90
" per condizioni morbose in atto	91
" per improvviso malore	92
" per sonno	93
" per ingestione di sostanze stupefacenti o psicotrope (art. 187)	94
Mancato uso di lenti correttive o apparecchi di protesi (art. 173)	95
Abbagliato	96
Per aver superato i periodi di guida prescritti (art.174)	97

CODICI ISTAT DEI RACCORDI AUTOSTRADALI E DELLE TANGENZIALI

R01 Raccordo Tangenziale Nord Città di Bologna (Crespellano-Aeroporto)
R02 Raccordo Autostradale Salerno-Avellino
R03 Raccordo Autostradale Siena-Firenze
R04 Raccordo Autostradale di Reggio Calabria
R05 Raccordo Autostradale Scalo Sicignano-Potenza
R06 Raccordo Autostradale Bettolle-Perugia
R07 Raccordo Autostradale Pavia-Autostrada A/7 (Milano-Serravalle) (Bereguardo-Pavia)
R08 Raccordo Autostradale Ferrara-Porto Garibaldi
R09 Raccordo Autostradale di Benevento
R10 Raccordo Autostradale Torino-Aeroporto di Caselle
R11 Raccordo Porto d'Ascoli-Ascoli Piceno
R12 Raccordo Autostradale Chieti-Pescara
R13 Raccordo Autostradale A/4-Trieste
R14 Raccordo Autostradale Trieste-Diramazione per Fernetti
R15 Tangenziale Ovest di Catania
R19 Raccordo La Spezia-Lerici
R28 Raccordo Autostradale Siena-Bettolle
R34 A 14 - Raccordo per Tangenziale di Bari
R36 Raccordo Molino Dorino (SS 011-SS 033)
R37 Raccordo Marco Polo (A 04-Aeroporto)
R38 Raccordo Autostradale Gazzada-Varese
R40 A 13 - Raccordo Padova Sud
R50 Grande Raccordo Anulare di Roma
R51 Roma-Fiumicino
R52 Bretella Aeroporto Falcone-Borsellino (Palermo-Punta Raisi)
R53 Raccordo A/5-SS 027 del Gran San Bernardo
R54 Raccordo Cimpello-Pian di Pan
R55 Raccordo Tolentino-Civitanova Marche
R56 Raccordo Tangenziale Nord Città di Bologna
R60 Raccordo A 01-Tangenziale Est di Milano
T01 Tangenziale Sud Torino
T02 Tangenziale Nord Torino
T03 Tangenziale Est-Ovest Napoli
T04 Tangenziale Pavia
T05 A 01 - Diramazione Capodichino
T06 A 06 - Diramazione per Fossano
T07 Tangenziale Est di Verona
T08 TR 01 - Traforo del Monte Bianco
T09 TR 02 - Traforo del Gran San Bernardo
T10 TR 04 - Traforo del Frejus
T11 Diramazione Roma Nord (Fiano-Roma)
T12 Diramazione Roma Sud (San Cesareo-Roma)
T13 Diramazione Moncalieri
T14 Diramazione Pinerolo
T15 Diramazione Abbadia
T20 Tangenziale sud di Verona
T21 Tangenziale sud di Brescia
T22 A 12 - Diramazione per Livorno

Figure 15 – ISTAT CTT/INC crash report form, codes.


RILEVAZIONE IST-00142

Istat
ISTITUTO NAZIONALE DI STATISTICA

ISTAT CTT/INC - Edizione 2019

RILEVAZIONE DEGLI INCIDENTI STRADALI CON LESIONI A PERSONE

DATA E LOCALITÀ DELL'INCIDENTE
ANNO MESE GIORNO
ORA MINUTI
PROVINCIA* Cod. Istat Provincia
COMUNE* Cod. Istat Comune
*L'elenco codici è disponibile sul sito www.istat.it (Strumenti/Definizioni e Classificazioni).

ORGANO DI RILEVAZIONE
Agente di Polizia Stradale 1
Carabiniere 2
Identificativo del Comando Staz. dei Carabinieri
Agente di Pubblica Sicurezza 3
Agente di Polizia Municipale o Locale 4
Altri 5
Agente di Polizia Provinciale 6

ORGANO COORDINATORE
Sezione Polizia Stradale 1
Gruppo Carabiniere 2
Uff. Comunale di Statistica dei Capoluoghi di Provincia:
Comune con oltre 250.000 abitanti 3
Altro capoluogo di Provincia 4

1. Localizzazione dell'incidente
(Specificare la denominazione della strada, numero, eventuale n° civico in forma chiara e leggibile)
NELL'ABITATO
Strada urbana 1
Provinciale entro l'abitato 2 SP N°
Statale entro l'abitato 3 SS N°
Regionale entro l'abitato 0 SR N°
FUORI ABITATO
Comunale extraurbana 4
Provinciale 5 SP N°
Statale 6 SS N°
Autostrada 7 N°
Altra strada 8
Regionale 9 SR N°
Indicare il codice Istat corrispondente ai raccordi autostradali e alle tangenziali, disponibile nella seconda pagina di copertina del presente blocco.
Progressiva chilometrica (indicare chilometri e metri)

TRONCO DI STRADA O AUTOSTRADA
diramazione; dir. A 1
dir. B; radd. 2
bis; dir. C 3
ter; bis dir. 4
quater; racc.; bis racc. 5
Autostrada carreggiata sinistra 6
Autostrada carreggiata destra 7
Autostrada svincolo entrata 8
Autostrada svincolo uscita 9
Autostrada svincolo tronco d.c. 10
Autostrada stazione 11
Altri casi 12

2. Luogo dell'incidente

TIPO DI STRADA
Una carr. senso unico 1
Una carr. doppio senso 2
Due carreggiate 3
Più di 2 carreggiate 4

PAVIMENTAZIONE
Strada pavimentata 1
Strada pavimentata dissestata 2
Strada non pavimentata 3

INTERSEZIONE
Incrocio 1
Rotatoria 2
Intersezione segnalata 3
Intersezione con semaforo o vigile 4
Intersezione non segnalata 5
Passaggio a livello 6

NON INTERSEZIONE
Rettilineo 7
Curva 8
Dosso, strettoia 9
Pendenza 10
Gall. illuminata 11
Gall. non illuminata 12

FONDO STRADALE
Asciutto 1
Bagnato 2
Sdrucciolevole 3
Ghiacciato 4
Innevato 5

SEGNALETICA
Assente 1
Verticale 2
Orizzontale 3
Verticale e orizzontale 4
Temporanea di cantiere 5

CONDIZIONI METEOROLOGICHE
Sereno 1
Nebbia 2
Pioggia 3
Grandine 4
Neve 5
Vento forte 6
Altro 7

3. Natura dell'incidente
A) TRA VEICOLI IN MARCIA
Scontro frontale 1
Scontro frontale-laterale 2
Scontro laterale 3
Tamponamento 4
B) TRA VEICOLO E PEDONI
Investimento di pedoni 5
C) VEICOLO IN MARCIA CHE URTA VEICOLO FERMO O ALTRO
Urto con veicolo in fermata o in arresto 6
Urto con veicolo in sosta 7
Urto con ostacolo 8
Urto con treno 9
D) VEICOLO IN MARCIA SENZA URTO
Fuoriuscita (sbandamento,...) 10
Infortunio per frenata improvvisa 11
Infortunio per caduta da veicolo 12

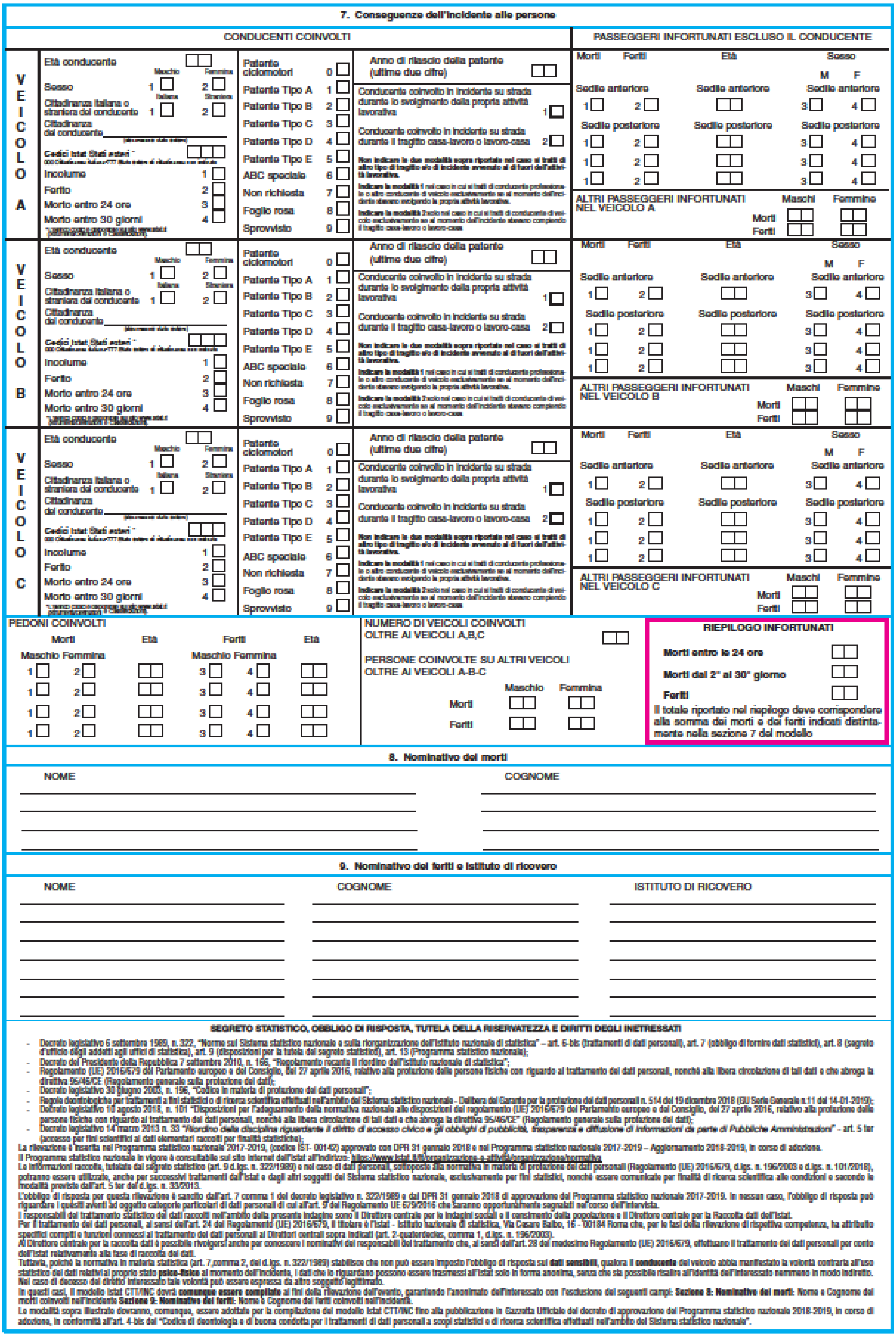
4. Tipo di veicoli coinvolti
VEICOLO: A B C
Autovettura privata 1
Autovettura privata con rimorchio 2
Autovettura pubblica 3
Autovettura di soccorso o di polizia 4
Autobus o filobus in servizio urbano 5
Autobus di linea o non di linea in extraurbana 6
Tram 7
Autocarro 8
Autotreno con rimorchio 9
Autoarticolato 10
Veicoli speciali 11
Trattore stradale o motrice 12
Macchina agricola 13
Velocipede 14
Ciclomotore 15
Motociclo a solo 16
Motociclo con passeggero 17
Motocarro o motofurgone 18
Veicolo a trazione animale o a braccia 19
Veicolo ignoto perché datosi alla fuga 20
Quadriciclo 21

CILINDRATA cc
Veic.A Veic.B Veic.C
PESO TOTALE A PIENO CARICO Q. li (SOLO VEICOLI TRASPORTO MERCI)
Veic.A Veic.B Veic.C

5. Circostanze presunte dell'incidente
Per inconvenienti di circolazione | Per difetti o avarie del veicolo | Per stato psico-fisico del conducente
Veicolo A
Veicolo B, Pedone od ostacolo
Indicare il codice Istat corrispondente alla circostanza presunta di incidente, disponibile nella seconda pagina di copertina del presente blocco

6. Veicoli coinvolti
Targa se veicolo nazionale
Veicolo A
Veicolo B
Veicolo C
Sigla se veicolo estero
ANNO DI PRIMA IMMATRICOLAZIONE (ultime due cifre)
Veicolo A
Veicolo B
Veicolo C

Figure 16 – ISTAT CTT/INC crash report form, crash general info section.


7. Conseguenze dell'incidente alle persone

CONDUCENTI COINVOLTI

PASSEGGERI INFORTUNATI ESCLUSO IL CONDUCENTE

VEICOLO A

Età conducente
Sesso: Maschio 1, Femmina 2
Cittadinanza italiana o straniera del conducente: Italiana 1, Straniera 2
Cittadinanza del conducente
Codici Istat Stati esteri
Incolume 1
Ferito 2
Morto entro 24 ore 3
Morto entro 30 giorni 4

Patente ciclomotori 0
Patente Tipo A 1
Patente Tipo B 2
Patente Tipo C 3
Patente Tipo D 4
Patente Tipo E 5
ABC speciale 6
Non richiesta 7
Foglio rosa 8
Sprovvisto 9

Anno di rilascio della patente (ultime due cifre)
Conducente coinvolto in incidente su strada durante lo svolgimento della propria attività lavorativa 1
Conducente coinvolto in incidente su strada durante il tragitto casa-lavoro o lavoro-casa 2
Non indicare le due modalità sopra riportate nel caso si tratti di altro tipo di tragitto e/o di incidente avvenuto al di fuori dell'attività lavorativa.

Morti, Feriti, Età, Sesso (M, F)
Sedile anteriore 1, 2; Sedile anteriore; Sedile anteriore 3, 4
Sedile posteriore 1, 2; Sedile posteriore; Sedile posteriore 3, 4

ALTRI PASSEGGERI INFORTUNATI NEL VEICOLO A
Maschi, Femmine
Morti
Feriti

VEICOLO B

Età conducente
Sesso: Maschio 1, Femmina 2
Cittadinanza italiana o straniera del conducente: Italiana 1, Straniera 2
Cittadinanza del conducente
Codici Istat Stati esteri
Incolume 1
Ferito 2
Morto entro 24 ore 3
Morto entro 30 giorni 4

Patente ciclomotori 0
Patente Tipo A 1
Patente Tipo B 2
Patente Tipo C 3
Patente Tipo D 4
Patente Tipo E 5
ABC speciale 6
Non richiesta 7
Foglio rosa 8
Sprovvisto 9

Anno di rilascio della patente (ultime due cifre)
Conducente coinvolto in incidente su strada durante lo svolgimento della propria attività lavorativa 1
Conducente coinvolto in incidente su strada durante il tragitto casa-lavoro o lavoro-casa 2
Non indicare le due modalità sopra riportate nel caso si tratti di altro tipo di tragitto e/o di incidente avvenuto al di fuori dell'attività lavorativa.

Morti, Feriti, Età, Sesso (M, F)
Sedile anteriore 1, 2; Sedile anteriore; Sedile anteriore 3, 4
Sedile posteriore 1, 2; Sedile posteriore; Sedile posteriore 3, 4

ALTRI PASSEGGERI INFORTUNATI NEL VEICOLO B
Maschi, Femmine
Morti
Feriti

VEICOLO C

Età conducente
Sesso: Maschio 1, Femmina 2
Cittadinanza italiana o straniera del conducente: Italiana 1, Straniera 2
Cittadinanza del conducente
Codici Istat Stati esteri
Incolume 1
Ferito 2
Morto entro 24 ore 3
Morto entro 30 giorni 4

Patente ciclomotori 0
Patente Tipo A 1
Patente Tipo B 2
Patente Tipo C 3
Patente Tipo D 4
Patente Tipo E 5
ABC speciale 6
Non richiesta 7
Foglio rosa 8
Sprovvisto 9

Anno di rilascio della patente (ultime due cifre)
Conducente coinvolto in incidente su strada durante lo svolgimento della propria attività lavorativa 1
Conducente coinvolto in incidente su strada durante il tragitto casa-lavoro o lavoro-casa 2
Non indicare le due modalità sopra riportate nel caso si tratti di altro tipo di tragitto e/o di incidente avvenuto al di fuori dell'attività lavorativa.

Morti, Feriti, Età, Sesso (M, F)
Sedile anteriore 1, 2; Sedile anteriore; Sedile anteriore 3, 4
Sedile posteriore 1, 2; Sedile posteriore; Sedile posteriore 3, 4

ALTRI PASSEGGERI INFORTUNATI NEL VEICOLO C
Maschi, Femmine
Morti
Feriti

PEDONI COINVOLTI
Morti: Maschio 1, Femmina 2, Età
Feriti: Maschio 3, Femmina 4, Età

NUMERO DI VEICOLI COINVOLTI OLTRE AI VEICOLI A,B,C

PERSONE COINVOLTE SU ALTRI VEICOLI OLTRE AI VEICOLI A-B-C
Maschio, Femmina
Morti
Feriti

RIEPILOGO INFORTUNATI
Morti entro le 24 ore
Morti dal 2° al 30° giorno
Feriti
Il totale riportato nel riepilogo deve corrispondere alla somma dei morti e dei feriti indicati distintamente nella sezione 7 del modello

8. Nominativo dei morti
NOME, COGNOME

9. Nominativo dei feriti e istituto di ricovero
NOME, COGNOME, ISTITUTO DI RICOVERO

SEGRETO STATISTICO, OBBLIGO DI RISPOSTA, TUTELA DELLA RISERVATEZZA E DIRITTI DEGLI INTERESSATI

- Decreto legislativo 6 settembre 1989, n. 322, "Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto nazionale di statistica" – art. 6-bis (trattamenti di dati personali), art. 7 (obbligo di fornire dati statistici), art. 8 (segreto d'ufficio degli addetti agli uffici di statistica), art. 9 (disposizioni per la tutela del segreto statistico), art. 13 (Programma statistico nazionale);
- Decreto del Presidente della Repubblica 7 settembre 2010, n. 166, "Regolamento recante il riordino dell'Istituto nazionale di statistica";
- Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati);
- Decreto legislativo 30 giugno 2003, n. 196, "Codice in materia di protezione dei dati personali";
- Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale - Delibera del Garante per la protezione dei dati personali n. 514 del 19 dicembre 2018 (GU Serie Generale n.11 del 14-01-2019);
- Decreto legislativo 10 agosto 2018, n. 101 "Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE" (Regolamento generale sulla protezione dei dati);
- Decreto legislativo 14 marzo 2013 n. 33 *"Riordino della disciplina riguardante il diritto di accesso civico e gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte di Pubbliche Amministrazioni"* - art. 5 ter (accesso per fini scientifici ai dati elementari raccolti per finalità statistiche);

La rilevazione è inserita nel Programma statistico nazionale 2017-2019, (codice IST- 00142) approvato con DPR 31 gennaio 2018 e nel Programma statistico nazionale 2017-2019 – Aggiornamento 2018-2019, in corso di adozione.
Il Programma statistico nazionale in vigore è consultabile sul sito internet dell'Istat all'indirizzo: https://www.istat.it/it/organizzazione-e-attività/organizzazione/normativa
Le informazioni raccolte, tutelate dal segreto statistico (art. 9 d.lgs. n. 322/1989) e nel caso di dati personali, sottoposte alla normativa in materia di protezione dei dati personali (Regolamento (UE) 2016/679, d.lgs. n. 196/2003 e d.lgs. n. 101/2018), potranno essere utilizzate, anche per successivi trattamenti dall'Istat e dagli altri soggetti del Sistema statistico nazionale, esclusivamente per fini statistici, nonché essere comunicate per finalità di ricerca scientifica alle condizioni e secondo le modalità previste dall'art. 5 ter del d.lgs. n. 33/2013.
L'obbligo di risposta per questa rilevazione è sancito dall'art. 7 comma 1 del decreto legislativo n. 322/1989 e dal DPR 31 gennaio 2018 di approvazione del Programma statistico nazionale 2017-2019. In nessun caso, l'obbligo di risposta può riguardare i quesiti aventi ad oggetto categorie particolari di dati personali di cui all'art. 9 del Regolamento UE 679/2016 che saranno opportunamente segnalati nel corso dell'intervista.
I responsabili del trattamento statistico dei dati raccolti nell'ambito della presente indagine sono il Direttore centrale per le indagini sociali e il censimento della popolazione e il Direttore centrale per la Raccolta dati dell'Istat.
Per il trattamento dei dati personali, ai sensi dell'art. 24 del Regolamento (UE) 2016/679, il titolare è l'Istat - Istituto nazionale di statistica, Via Cesare Balbo, 16 - 00184 Roma che, per le fasi della rilevazione di rispettiva competenza, ha attribuito specifici compiti e funzioni connessi al trattamento dei dati personali ai Direttori centrali sopra indicati (art. 2-quaterdecies, comma 1, d.lgs. n. 196/2003).
Ai Direttore centrale per la raccolta dati è possibile rivolgersi anche per conoscere i nominativi dei responsabili del trattamento che, ai sensi dell'art. 28 del medesimo Regolamento (UE) 2016/679, effettuano il trattamento dei dati personali per conto dell'Istat relativamente alla fase di raccolta dei dati.
Tuttavia, poiché la normativa in materia statistica (art. 7,comma 2, del d.lgs. n.322/1989) stabilisce che non può essere imposto l'obbligo di risposta sui **dati sensibili**, qualora il **conducente** del veicolo abbia manifestato la volontà contraria all'uso statistico dei dati relativi al proprio stato **psico-fisico** al momento dell'incidente, i dati che lo riguardano possono essere trasmessi all'Istat solo in forma anonima, senza che sia possibile risalire all'identità dell'interessato nemmeno in modo indiretto. Nel caso di decesso del diritto interessato tale volontà può essere espressa da altro soggetto legittimato.
In questi casi, il modello Istat CTT/INC dovrà **comunque essere compilato** ai fini della rilevazione dell'evento, garantendo l'anonimato dell'interessato con l'esclusione dei seguenti campi: **Sezione 8: Nominativo dei morti:** Nome e Cognome dei morti coinvolti nell'incidente **Sezione 9: Nominativo dei feriti:** Nome e Cognome dei feriti coinvolti nell'incidente.
Le modalità sopra illustrate dovranno, comunque, essere adottate per la compilazione del modello Istat CTT/INC fino alla pubblicazione in Gazzetta Ufficiale del decreto di approvazione del Programma statistico nazionale 2018-2019, in corso di adozione, in conformità all'art. 4-bis del "Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale".

Figure 17 – ISTAT CTT/INC crash report form, involved vehicles and driver/passengers data section.

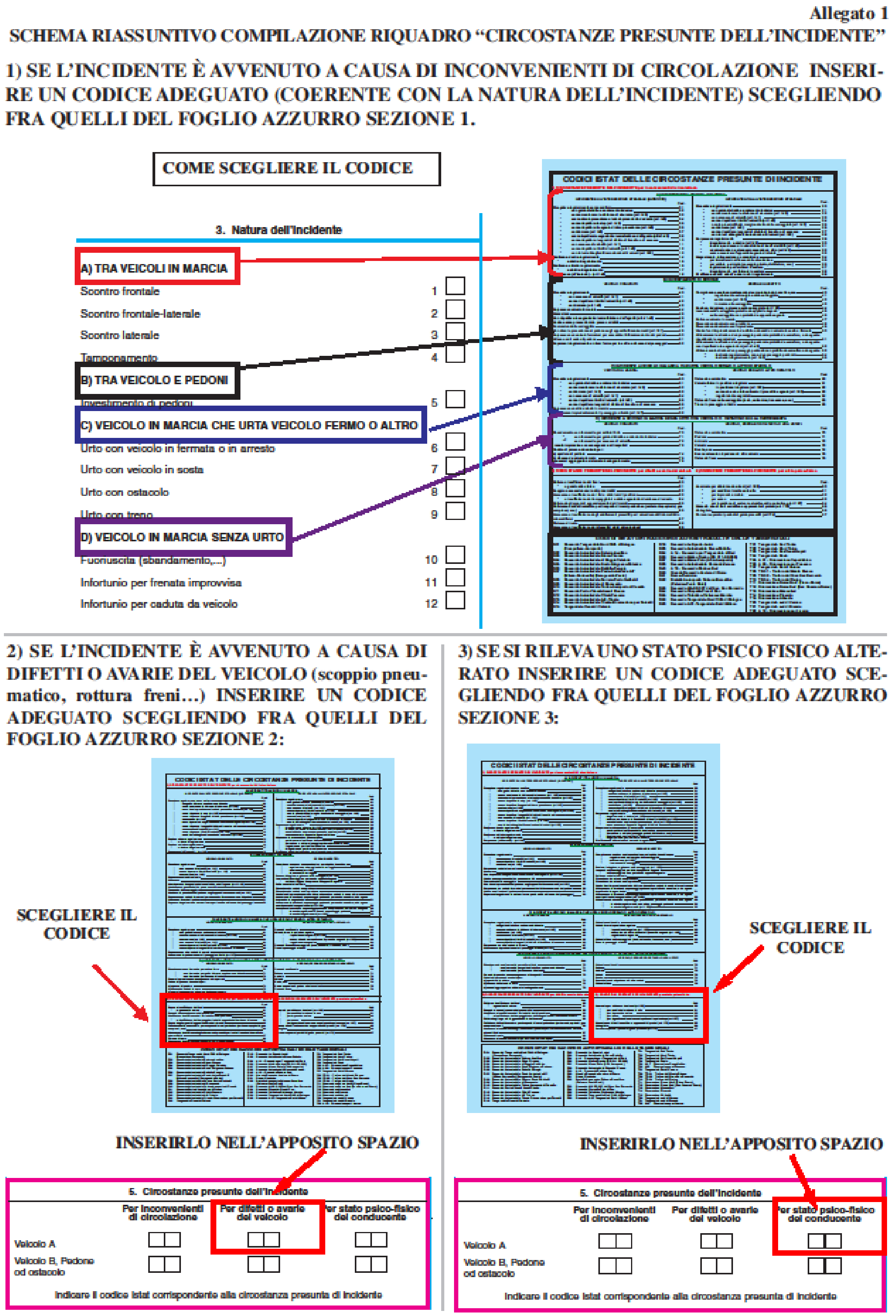


Figure 18 – ISTAT CTT/INC crash report form, instructions to fill in the form.

Crashes are classified through 118 variables that can be grouped into four macro-areas: 1) Crash characteristics which include temporal information, location, nature and presumed circumstances of crashes; 2) Road and environment area includes road characteristics and environmental conditions; 3) Traffic units include vehicle characteristics; 4) People involved section includes the characteristics of the occupants, broken down by passengers, drivers and pedestrians. Further 6 variables regarding detailed crash information as well as driver and pedestrian psychophysical states were used in the study (provided by ISTAT for research support). Several categories were aggregated and recoded to avoid extremely small occurrences of some categories, to remove redundant information and, finally, make the models easier to interpret.

Italian data used in this research refer to the five-year period 2014-2018. The dataset was rearranged in 20 categorical variables and consists of 874’847 crashes occurred on national roads. As the aim of this research is to study factors influencing the severity of pedestrian crashes, only crashes were at least one pedestrian was involved was considered eligible for the study. Thus, the final version of the crash dataset contains 101'032 pedestrian crashes only (representing 11.55% of all crashes). The Italian crash database makes no distinction between slight injury and serious injury, gathering crash severity with two different categories: injury crashes (n= 98'063; 97.06% of the total pedestrian crashes) and fatal crashes (n= 2'969; 2.94% of the total pedestrian crashes). The severity of the crashes is classified according to the severity of the most seriously injured person involved in the crash. It is considered fatal a crash where at least one person is killed instantly or within the thirtieth day beginning on the day in which the crash occurred.

The 20 variable considered in the research were: 1) Day of Week, 2) Season, 3) Road type, 4) Area, 5) Lighting, 6) Alignment,7) Pavement, 8) Weather, 10) Vehicle type, 11) Vehicle age, 12) Vehicle Defect, 13) Driver behaviour, 14) Driver psychophysical state 15) Pedestrian age, and 16) Pedestrian gender, 17) Pedestrian psychophysical state, 18) Pedestrian behaviour, 19) Pedestrian age, and 20) Pedestrian gender.

The variables “Day of week” and “Season” were generated by combining the information related to time data (year, month, and day). The variable “Lighting” was generated using the date of crash, the time of crash occurrence, and the geographical coordinates of the place. Then, the exact time of dawn and dusk was assessed using the open source software R and the R package “SUNCALC” and the variable was finally classified as a binary variable: day and night. The variable “Vehicle Age” has been created by using the year of the vehicle registration.

Below, the Italian pedestrian crash database is reported in Table 37 - Table 39.

Table 37 – Italian descriptive statistics related to pedestrian data (part A).

Variable	Total	%	Fatal	
			N	%
Total	101'032	100.00	2'969	2.94
Area				
Rural	4'878	4.83	646	13.24
Urban	96'154	95.17	2'323	2.42
Road Type				
Motorway	330	0.33	91	27.58
Rural national	1'059	1.05	175	16.53
Rural provincial	1'851	1.83	259	13.99
Rural Municipal	1'638	1.62	121	7.39
Urban national	3'163	3.13	174	5.50
Urban provincial	4'968	4.92	336	6.76
Urban Municipal	88'023	87.12	1'813	2.06
Alignment				
Curve	4'377	4.33	190	4.34
Unsignalised Intersection	23'398	23.16	438	1.87
Roundabout	2'141	2.12	28	1.31
Signalized Intersection	6'282	6.22	102	1.62
Tangent	63'334	62.69	2'145	3.39
Tunnel	102	0.10	5	4.90
Other	1'398	1.38	61	4.36
Day of Week				
Weekday	80'030	79.21	2'162	2.70
Weekend	21'002	20.79	807	3.84
Season				
Autumn	35'909	35.54	1'094	3.05
Spring	23'525	23.28	572	2.43
Summer	14'928	14.78	495	3.32
Winter	26'670	26.40	808	3.03
Lighting				
Day	70'903	70.18	1'526	2.15
Night	30'129	29.82	1'443	4.79
Pavement				
Dry	83'117	82.27	2'466	2.97
Slippery	236	0.23	8	3.39
Snowy/Frozen	254	0.25	5	1.97
Wet	17'425	17.25	490	2.81
Weather				
Clear	82'796	81.95	2'459	2.97
Fog	689	0.68	25	3.63
High winds	100	0.10	2	2.00
Raining	11'974	11.85	308	2.57
Snowing	236	0.23	4	1.69
Other	5'237	5.18	171	3.27
Crash Severity				
Fatal	2'969	2.94	2'969	100.00
Injury	98'063	97.06	-	0.00

Table 38 – Italian descriptive statistics related to pedestrian data (part B).

Variable	Total	%	Fatal	
			N	%
Vehicle Type				
Bicycle	1'837	1.82	17	0.93
Car	76'390	75.61	2'189	2.87
PTW	12'192	12.07	259	2.12
Truck	7'004	6.93	418	5.97
Other	3'609	3.57	86	2.38
Vehicle Age				
0-10	49'600	49.09	1'443	2.91
10-20	20'888	20.67	659	3.15
>20	2'781	2.75	97	3.49
Missing	25'985	25.72	753	2.90
Not applied	1'778	1.76	17	0.96
Vehicle Defect				
Defect	306	0.30	23	7.52
No defect	100'726	99.70	2'946	2.92
Driver Behaviour				
Disobeying pedestrian crossing facility	35'563	35.20	814	2.29
Disobeying stop sign	131	0.13	11	8.40
Distraction	808	0.80	47	5.82
Illegal travel direction	740	0.73	19	2.57
Manoeuvring	10'904	10.79	204	1.87
Normal	26'096	25.83	886	3.40
Speeding	9'416	9.32	545	5.79
Tailgating	731	0.72	32	4.38
Other	16'643	16.47	411	2.47
Driver Psychophysical State				
Alcohol	815	0.81	83	10.18
Dazzled	662	0.66	39	5.89
Drug	230	0.23	54	23.48
Exceeding the prescribed driving period	9	0.01	2	22.22
Illness	109	0.11	14	12.84
Normal	99'142	98.13	2'767	2.79
Sleeping	49	0.05	8	16.33
Uncorrected, defective eyesight	16	0.02	2	12.50
Driver Age				
0-17	1'282	1.27	22	1.72
18-24	8'474	8.39	354	4.18
25-44	30'389	30.08	1'023	3.37
45-54	20'074	19.87	579	2.88
55-64	14'238	14.09	407	2.86
65-74	10'710	10.60	241	2.25
75+	9'645	9.55	252	2.61
Missing	6'220	6.16	91	1.46
Driver Gender				
Female	24'467	24.22	472	1.93
Male	73'850	73.10	2'461	3.33
Missing	2'715	2.69	36	1.33

Table 39 – Italian descriptive statistics related to pedestrian data.

Variable	Total	%	Fatal	
			N	%
Pedestrian Behaviour				
Crossing on pedestrian crossing facility	38'180	37.79	859	2.25
Crossing outside pedestrian crossing facility	24'106	23.86	901	3.74
Walking facing the traffic	790	0.78	36	4.56
Walking back to the traffic	5'571	5.51	224	4.02
Walking Regularly	8'478	8.39	266	3.14
Other	23'907	23.66	683	2.86
Pedestrian Psychophysical State				
Alcohol	307	0.30	20	6.51
Drug	38	0.04	2	5.26
Illness	94	0.09	11	11.70
Normal	100'583	99.56	2'936	2.92
Sleeping	10	0.01	0	0.00
Pedestrian Age				
0-14	9'426	9.33	86	0.91
15-24	10'967	10.85	135	1.23
25-44	19'914	19.71	328	1.65
45-54	14'798	14.65	279	1.89
55-64	12'873	12.74	346	2.69
65-74	12'398	12.27	453	3.65
75+	19'484	19.28	1'315	6.75
Missing	1'172	1.16	27	2.30
Pedestrian Gender				
Female	53'840	53.29	1'083	2.01
Male	47'192	46.71	1'886	4.00

CHAPTER V ~ RESULTS

In this chapter the main results obtained by each method and for each case study were reported. In-depth results were provided at the end of this dissertation thesis in the appendix relative to each case study.

5.1 Great Britain results

All the explanatory variables reported in the descriptive statistics (from Table 28 to Table 33) were tested for inclusion in the econometric models. The estimation results are reported in Table 40 and Table 41 for the multinomial logit, in Table 44 and Table 45 are provided the results of the mixed multinomial logit, in Table 48 and Table 49 are presented the results related to the ordered logit, the in Table 52 and Table 53 are reported the results of the mixed ordered logit. Regarding the machine learning tools, Figure 19 presents the classification tree, Table 60 and Table 61 provide the variable importance for fatal and serious injury classifications in RF, Table 64 and Table 65 present partially the results of AR, Table 68 provides the variable importance for fatal and serious injury classifications in SVM, and Table 71 provides summary results for ANN tool.

Furthermore, the confusion matrix and all the performance metrics evaluated are reported for each method.

5.1.1 Multinomial logit

Statistically significant explanatory variables were 20 and significant indicator variables associated with these categorical variables were 41. 36 significant indicators described the fatal crashes and 35 significant indicators described the serious injury crashes. Model's McFadden Pseudo R^2 is equal to 0.16, which indicates a good fit. The most influential variable was the pedestrian age. Compared to young pedestrians (35-44), the elderly pedestrians (aged 75 or more) increased the probability of fatal crashes with an OR of 13.17. Another significant indicator is the speed limit ≥ 50 mph. the indicator exhibited an OR equal to 9.27.

Table 40 – Multinomial logit: parameter estimates and goodness of fit measures, Great Britain (Part A).

Variable	Fatal				Serious			
	Estimate	OR	Std. Err.	P>|z|	Estimate	OR	Std. Err.	P>|z|
Intercept	-5.215	0.005	0.129	<0.001	-1.529	0.217	0.031	<0.001
Number of vehicles (1 vehicle as baseline)								
2	0.682	1.978	0.106	<0.001	0.183	1.201	0.042	<0.001
≥ 3	1.170	3.222	0.187	<0.001	0.498	1.645	0.091	<0.001
1st Road class (C as baseline)								
B					0.091	1.095	0.031	0.004
A	0.558	1.747	0.067	<0.001	0.095	1.100	0.022	<0.001
Motorway	0.979	2.662	0.263	<0.001	0.484	1.623	0.230	0.035
Speed Limit (20 mph as baseline)								
30	0.382	1.465	0.125	0.002	0.073	1.076	0.037	0.044
40	1.384	3.991	0.163	<0.001	0.565	1.759	0.057	<0.001
≥ 50	2.227	9.272	0.164	<0.001	0.638	1.893	0.064	<0.001
Area (Urban as baseline)								
Rural	0.347	1.415	0.086	<0.001				
Junction detail (T or staggered junction as baseline)								
Not at junction					-0.034	0.967	0.015	0.021
Roundabout	-0.353	0.703	0.187	0.059	-0.082	0.921	0.048	0.091
Pedestrian crossing human control (None within 50 metres as baseline)								
School crossing patrol					-0.204	0.815	0.120	0.089
Pedestrian crossing physical facilities (None within 50 metres as baseline)								
Zebra	-0.743	0.476	0.169	<0.001	-0.212	0.809	0.037	<0.001
Pelican	0.254	1.289	0.094	0.007	0.114	1.121	0.033	0.001
Lighting (Daylight as baseline)								
Darkness	1.090	2.974	0.066	<0.001	0.290	1.336	0.022	<0.001
Pavement (Dry as baseline)								
Wet or damp	0.142	1.153	0.078	0.069	0.049	1.050	0.027	0.075
Snow	-0.877	0.416	0.306	0.004				
Day of week (Weekday as baseline)								
Weekend	0.356	1.428	0.066	<0.001	0.126	1.134	0.023	<0.001

Table 41 - Multinomial logit: parameter estimates and goodness of fit measures, Great Britain (Part B).

Variable	Fatal				Serious			
	Estimate	OR	Std. Err.	P>|z|	Estimate	OR	Std. Err.	P>|z|
Vehicle type (Car as baseline)								
Bicycle	-1.290	0.275	0.366	<0.001	0.141	1.151	0.064	0.028
Bus	0.710	2.034	0.164	<0.001				
PTW<500	-1.122	0.326	0.224	<0.001	-0.103	0.902	0.051	0.044
Truck	1.515	4.549	0.124	<0.001				
Vehicle towing and articulation (No towing/articulation as baseline)								
Articulated vehicle	1.228	3.414	0.221	<0.001	0.855	2.351	0.141	<0.001
Vehicle propulsion code (Petrol as baseline)								
Heavy oil vehicles	0.284	1.328	0.072	<0.001	0.170	1.185	0.033	<0.001
Hybrid vehicles	-0.466	0.628	0.283	0.100	-0.289	0.749	0.062	<0.001
Vehicle age (≤ 15 years as baseline)								
>15	0.327	1.387	0.128	0.011	0.213	1.237	0.043	<0.001
Vehicle manoeuvre (Moving off as baseline)								
Going ahead	1.126	3.083	0.073	<0.001	0.505	1.657	0.026	<0.001
Turning manoeuvre					0.140	1.150	0.035	<0.001
Reversing manoeuvre					-0.152	0.859	0.044	0.001
Vehicle skidding and overturning (No as baseline)								
Yes	1.165	3.206	0.117	<0.001	0.480	1.616	0.056	<0.001
Driver gender (Male as baseline)								
Female	-0.293	0.746	0.078	<0.001				
Driver age (35-44 as baseline)								
≤ 24	0.596	1.815	0.091	<0.001	0.272	1.313	0.030	<0.001
25-34	0.293	1.340	0.076	<0.001	0.145	1.156	0.024	<0.001
Pedestrian gender (Male as baseline)								
Female	-0.155	0.856	0.064	0.015	-0.072	0.931	0.019	<0.001
Pedestrian age (35-44 as baseline)								
0-14	-0.837	0.433	0.137	<0.001				
15-24	-0.534	0.586	0.105	<0.001				
25-34	-0.303	0.739	0.103	0.003				
45-54					0.154	1.166	0.031	<0.001
55-64	0.633	1.883	0.110	<0.001	0.417	1.517	0.033	<0.001
65-74	1.295	3.651	0.111	<0.001	0.770	2.160	0.035	<0.001
≥75	2.578	13.171	0.092	<0.001	1.111	3.037	0.034	<0.001
Log likelihood null model			-48,217.27					
Log likelihood full model			-40,469.52					
R^2McFadden			0.161					

Overall, the MNL model exhibited 93% of correct classification for slight injury crashes, 14% for serious injury, and 25% for fatal crashes with a global accuracy superior to 70%. However, the total correct classification was also evaluated considering the model performances exhibited for the classification of slight injuries, the most frequent class.

Table 42 – Confusion matrix for the multinomial logit, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	46,329	2,966	336
	Serious	13,656	2,316	387
	Fatal	569	453	344

Table 43 – Performance of metrics for the multinomial logit model, Great Britain.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.932	0.985	0.362
TP_{rate} (Acc+)	0.142	0.252	0.727
Precision	0.404	0.322	0.668
F-measure	0.210	0.283	0.676
G-mean	0.363	0.498	0.381
AUC	0.621	0.871	0.650
Acc	0.727		
Err	0.273		

5.1.2 Mixed multinomial logit

Results for both fixed and random variables are reported in Table 44 and Table 45. The log-likelihood at zero (-48,217) and at convergence (-39,565) gives a McFadden R^2 of 0.18 which is an acceptable result. It is also the highest value exhibited among the parametric models in the British context. The goodness of fit results and LR test results show that the random model provides a significant improvement compared to the fixed parameters model. The $\chi 2$ of the LR test is 1,808.11 with 3 degrees of freedom and p-value <0.001, showing that the RPMNL model is superior to the standard MNL model with over 99.9% of confidence. Three indicator variables showed normally distributed random parameters, with statistically significant standard deviations indicating significant unobserved heterogeneity in the data. These variables are (1) going ahead vehicle manoeuvres (fatal), (2) roundabout (fatal), and (3) pedestrian age greater or equal to 75 (serious injury). In the prediction of fatal severity, the indicator variable roundabout showed a normal distribution with a mean of -2.477 and a standard deviation of 2.583. This means that for 83.1% at roundabouts the probability of the fatal outcome decreased while for 16.9% of the observations the probability of a fatal outcome increased. Similarly, the indicator variable going ahead vehicle manoeuvre showed a normal distribution with a mean of 0.831 and a standard deviation of 0.997. This means that for 79.8% of the observations with vehicles that manoeuvred going ahead the probability of the fatal outcome increased while for 20.2% of the observations the probability of a fatal outcome decreased. In the prediction of severe injury, the indicator variable pedestrian age ≥ 75 showed a normal distribution with a mean of 0.297 and a standard deviation of 3.852. This means that for 53.1% of the observations with pedestrian age ≥ 75 the probability of severe injury increased while for 46.9% of the observations the probability of severe injury decreased. The fixed coefficients of the random parameter multinomial logit were similar in sign and magnitude to the standard multinomial model.

Table 44 – Mixed multinomial logit: parameter estimates and goodness of fit measures, Great Britain (Part A).

Variable	Fatal				Serious			
	Estimate	OR	Std. Err.	P>|z|	Estimate	OR	Std. Err.	P>|z|
Intercept	-5.364	0.005	0.196	<0.001	-1.041	0.353	0.043	<0.001
Number of vehicles (1 vehicle as baseline)								
2	0.735	2.085	0.117	<0.001	0.175	1.191	0.042	<0.001
≥ 3	1.218	3.380	0.199	<0.001	0.493	1.637	0.090	<0.001
1st Road class(C as baseline)								
B					0.108	1.114	0.032	0.001
A	0.577	1.781	0.072	<0.001	0.104	1.110	0.022	<0.001
Motorway	1.043	2.838	0.284	<0.001	0.448	1.565	0.215	0.037
Speed Limit (20 mph as baseline)								
30	0.423	1.527	0.137	0.002	0.051	1.052	0.030	0.088
40	1.478	4.384	0.178	<0.001	0.524	1.689	0.055	<0.001
≥ 50	2.431	11.370	0.186	<0.001	0.582	1.790	0.061	<0.001
Area (Urban as baseline)								
Rural	0.377	1.458	0.096	<0.001				
Junction detail (T or staggered junction as baseline)								
Not at junction					-0.044	0.957	0.021	0.035
Roundabout	**-2.477**	**0.084**	**0.966**	**0.010**	-0.107	0.899	0.059	0.069
Pedestrian crossing human control (None within 50 metres as baseline)								
School crossing patrol					-0.207	0.813	0.123	0.093
Pedestrian crossing physical facilities (None within 50 metres as baseline)								
Zebra	-0.781	0.458	0.188	<0.001	-0.231	0.794	0.039	<0.001
Pelican	0.280	1.323	0.098	0.004	0.103	1.108	0.030	0.001
Lighting (Daylight as baseline)								
Darkness	1.164	3.203	0.076	<0.001	0.289	1.335	0.022	<0.001
Pavement (Dry as baseline)								
Wet or damp	0.153	1.165	0.075	0.041	0.040	1.041	0.023	0.078
Snow	-1.045	0.352	0.359	0.004				
Day of week (Weekday as baseline)								
Weekend	0.373	1.452	0.074	<0.001	0.123	1.131	0.023	<0.001

Table 45 – Mixed multinomial logit: parameter estimates and goodness of fit measures, Great Britain (Part B).

Variable	**Fatal**				**Serious**			
	Estimate	**OR**	**Std. Err.**	**P>|z|**	**Estimate**	**OR**	**Std. Err.**	**P>|z|**
Vehicle type (Car as baseline)								
Bicycle	-1.427	0.240	0.403	<0.001	0.223	1.250	0.067	0.001
Bus	0.634	1.885	0.147	<0.001				
PTW<500	-1.288	0.276	0.254	<0.001	-0.112	0.894	0.053	0.033
Truck	1.674	5.333	0.151	<0.001				
Vehicle towing and articulation (No towing/articulation as baseline)								
Yes	1.272	3.568	0.234	<0.001	0.833	2.300	0.141	<0.001
Vehicle propulsion code (Petrol as baseline)								
Heavy oil vehicles	0.284	1.328	0.072	<0.001	0.170	1.185	0.033	<0.001
Hybrid vehicles	-0.466	0.628	0.283	0.100	-0.289	0.749	0.062	<0.001
Vehicle age(≤ 15 years as baseline)								
>15	0.317	1.373	0.086	<0.001	0.153	1.165	0.023	<0.001
Vehicle manoeuvre (Moving off as baseline)								
Going ahead	**0.831**	**2.296**	**0.154**	**<0.001**	0.513	1.670	0.027	<0.001
Turning manoeuvre					0.143	1.154	0.037	<0.001
Reversing manoeuvre					-0.255	0.775	0.051	<0.001
Vehicle skidding and overturning(No as baseline)								
Yes	1.266	3.547	0.133	<0.001	0.450	1.568	0.054	<0.001
Driver gender (Male as baseline)								
Female	-0.343	0.710	0.092	<0.001				
Driver age (35-44 as baseline)								
≤ 24	0.635	1.887	0.101	<0.001	0.294	1.342	0.031	<0.001
25-34	0.336	1.399	0.084	<0.001	0.152	1.164	0.025	<0.001
Pedestrian gender (Male as baseline)								
Female	-0.156	0.856	0.070	0.027	-0.097	0.908	0.020	<0.001
Pedestrian age (35-44 as baseline)								
0-14	-0.884	0.413	0.148	<0.001				
15-24	-0.592	0.553	0.116	<0.001				
25-34	-0.342	0.710	0.114	0.003				
45-54					0.157	1.170	0.031	<0.001
55-64	0.668	1.950	0.118	<0.001	0.426	1.531	0.033	<0.001
65-74	1.367	3.924	0.120	<0.001	0.785	2.192	0.035	<0.001
≥75	2.279	9.767	0.223	<0.001	**0.297**	**1.346**	**0.179**	**0.097**
Standard deviation of random parameter								
Going ahead vehicle manoeuvre	**0.997**	**2.710**	**0.195**	**<0.001**				
Roundabout	**2.583**	**13.237**	**0.643**	**<0.001**				
Pedestrian age ≥75					**3.853**	**47.134**	**1.036**	**<0.001**
Log likelihood null model			-48,217.27					
Log likelihood full model			-39,565.47					
R^2McFadden			0.179					

The RPMNL model exhibited 82% of correct classification for slight injury crashes, 40% for serious injury, and 42% for fatal crashes with a global accuracy superior to 71%. As for MNL model, the total correct classification was also evaluated considering the model performances exhibited for the classification of slight injuries, the most frequent class. It is noteworthy to observe that even if the overall accuracy of RPMNL model slightly differs from that of MNL, the performance of serious injury and fatal crashes correctly classified improved.

Table 46 – Confusion matrix for the mixed multinomial logit, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	40,778	8,714	139
	Serious	9,669	6,606	84
	Fatal	332	462	572

Table 47 – Performance of metrics for the mixed multinomial logit model, Great Britain.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.818	0.995	0.527
TP_{rate} (Acc+)	0.404	0.419	0.712
Precision	0.419	0.719	0.708
F-measure	0.411	0.529	0.709
G-mean	0.575	0.646	0.584
AUC	0.683	0.937	0.699
Acc	0.712		
Err	0.288		

5.1.3 Ordered logit

The ordered logit model was carried out to capture the ordinal nature of the response variable. A positive (or negative) parameter implied the likelihood (or unlikelihood) of a severe injury with an increasing value of the explanatory variable and a reduction in the likelihood of a slight injury (Naik et al., 2016). Statistically significant explanatory variables were 18 and significant indicator variables associated with these categorical variables were 35 (see Table 48 and Table 49).

Table 48 – Ordered logit: parameter estimates and goodness of fit measures, Great Britain (Part A).

Variable	**Estimate**	**OR**	**Std. Err.**	**P>|z|**
Number of vehicles (1 vehicle as baseline)				
2	0.262	1.300	0.039	<0.001
≥ 3	0.613	1.846	0.083	<0.001
1st road class (C as baseline)				
B	0.108	1.114	0.030	<0.001
A	0.172	1.188	0.021	<0.001
Motorway	1.003	2.726	0.184	<0.001
Speed Limit (20 mph as baseline)				
30	0.076	1.079	0.029	0.008
40	0.615	1.850	0.051	<0.001
≥ 50	1.079	2.942	0.056	<0.001
Junction detail (T or staggered junction as baseline)				
Not at junction	-0.046	0.955	0.020	0.021
Roundabout	-0.099	0.906	0.055	0.071
Pedestrian crossing human control (None within 50 metres as baseline)				
School crossing patrol	-0.244	0.783	0.120	0.042
Pedestrian crossing physical facilities (None within 50 metres as baseline)				
Zebra	-0.226	0.798	0.037	<0.001
Pelican	0.103	1.108	0.028	<0.001
Lighting (Daylight as baseline)				
Darkness	0.409	1.505	0.021	<0.001
Pavement (Dry as baseline)				
Wet or damp	0.047	1.048	0.022	0.035
Snow	-0.236	0.790	0.091	0.010
Day of week (Weekday as baseline)				
Weekend	0.150	1.162	0.022	<0.001

Table 49 – Ordered logit: parameter estimates and goodness of fit measures, Great Britain (Part B).

Variable	Estimate	OR	Std. Err.	P>|z|
Vehicle type (Car as baseline)				
Bus	0.184	1.202	0.046	<0.001
PTW<500	-0.158	0.854	0.051	0.002
Truck	0.462	1.587	0.066	<0.001
Vehicle towing and articulation (No towing/articulation as baseline)				
Yes	1.260	3.525	0.129	<0.001
Vehicle propulsion code (Petrol as baseline)				
Heavy oil vehicles	0.119	1.126	0.022	<0.001
Hybrid vehicles	-0.340	0.712	0.071	<0.001
Vehicle age (≤ 15 years as baseline)				
>15	0.232	1.261	0.042	<0.001
Vehicle manoeuvre (Moving off as baseline)				
Going ahead	0.587	1.799	0.023	<0.001
Turning manoeuvre	0.187	1.206	0.032	<0.001
Vehicle skidding and overturning				
Yes	0.607	1.835	0.051	<0.001
Driver age (35-44 as baseline)				
≤ 24	0.304	1.355	0.029	<0.001
25-34	0.155	1.168	0.024	<0.001
Pedestrian gender (Male as baseline)				
Female	-0.080	0.923	0.019	<0.001
Pedestrian age (35-44 as baseline)				
0-14	-0.171	0.843	0.025	<0.001
45-54	0.233	1.262	0.031	<0.001
55-64	0.516	1.675	0.033	<0.001
65-74	0.895	2.447	0.035	<0.001
≥75	1.393	4.027	0.033	<0.001
Cut points				
Cut1	2.381		0.039	
Cut2	5.385		0.049	
Log likelihood null model	-48,217.27			
Log likelihood full model	-41,017.93			
R^2McFadden	0.149			

Overall, the OL model exhibited 31% of correct classification for slight injury crashes, 82% for serious injury, and 0% for fatal crashes with a global accuracy superior to 42%. The OL model, even considering the most frequent class to assess the total correct classification the model global accuracy was very low.

Table 50 – Confusion matrix for the ordered logit, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	15,201	34,430	0
	Serious	2,976	13,383	0
	Fatal	36	1,328	2

Table 51 – Performance of metrics for the ordered logit model, Great Britain.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.298	1.000	0.694
TP_{rate} (Acc+)	0.818	0.001	0.424
Precision	0.272	1.000	0.701
F-measure	0.409	0.003	0.430
G-mean	0.494	0.038	0.489
AUC	0.608	0.851	0.636
Acc	0.424		
Err	0.576		

5.1.4 Mixed ordered logit

Results for both fixed and random variables are reported in Table 52 and Table 53. The goodness of fit results and LR test results show that random model provides a significant improvement compared to the fixed parameters model. The χ^2 of the LR test is 1,832.61 with 1 degree of freedom and p-value <0.001, showing that the RPOL model is superior to the standard OL model with over 99.9% of confidence.

One indicator variable showed normally distributed random parameters, with statistically significant standard deviation indicating significant unobserved heterogeneity in the data (Table 53). This variable is pedestrian age greater or equal to 75. In the prediction of both fatal and severe injury severity, the indicator variable pedestrian age ≥ 75 showed a normal distribution with a mean of 0.258 and a standard deviation of 0.580. This means that for 67.8% of the observations with pedestrian age ≥ 75 the probability of the most severe injury increased while for 32.8% of the observations the probability decreased. Similarly to the unordered models, the fixed coefficients of the random parameter ordered logit were similar in sign and magnitude to the standard ordinal model.

Table 52 – Mixed ordered logit: parameter estimates and goodness of fit measures, Great Britain (Part A).

Variable	Estimate	OR	Std. Err.	P>|z|
Number of vehicles (1 vehicle as baseline)				
2	0.195	1.215	0.039	<0.001
≥ 3	0.571	1.770	0.083	<0.001
1st road class (C as baseline)				
B	0.110	1.116	0.030	0.001
A	0.150	1.162	0.021	<0.001
Motorway	0.925	2.522	0.184	<0.001
Speed Limit (20 mph as baseline)				
30	0.090	1.094	0.029	0.002
40	0.627	1.872	0.052	<0.001
≥ 50	1.029	2.798	0.061	<0.001
Junction detail (T or staggered junction as baseline)				
Not at junction	-0.057	0.945	0.020	0.004
Roundabout	-0.133	0.875	0.056	0.017
Pedestrian crossing human control (None within 50 metres as baseline)				
School crossing patrol	-0.274	0.760	0.121	0.024
Pedestrian crossing physical facilities (None within 50 metres as baseline)				
Zebra	-0.228	0.796	0.037	<0.001
Pelican	0.122	1.130	0.028	<0.001
Lighting (Daylight as baseline)				
Darkness	0.336	1.399	0.021	<0.001
Pavement (Dry as baseline)				
Wet or damp	0.071	1.074	0.022	0.001
Snow	-0.240	0.787	0.091	0.009
Day of week (Weekday as baseline)				
Weekend	0.133	1.142	0.022	<0.001

Table 53 – Mixed ordered logit: parameter estimates and goodness of fit measures, Great Britain (Part B).

Variable	Estimate	OR	Std. Err.	P>|z|
Vehicle type (Car as baseline)				
Bus	0.142	1.153	0.046	0.002
PTW<500	-0.149	0.862	0.051	0.004
Truck	0.424	1.528	0.066	<0.001
Vehicle towing and articulation(No towing/articulation as baseline)				
Yes	1.299	3.666	0.129	<0.001
Vehicle propulsion code (Petrol as baseline)				
Heavy oil vehicles	0.209	1.232	0.020	<0.001
Hybrid vehicles	-0.252	0.777	0.070	<0.001
Vehicle age (≤ 15 years as baseline)				
>15	0.237	1.267	0.042	<0.001
Vehicle manoeuvre (Moving off as baseline)				
Going ahead vehicle manoeuvre	0.536	1.709	0.025	<0.001
Turning manoeuvre	0.203	1.225	0.035	<0.001
Vehicle skidding and overturning				
Yes	0.593	1.809	0.051	<0.001
Driver age (35-44 as baseline)				
≤ 24	0.332	1.394	0.029	<0.001
25-34	0.171	1.186	0.023	<0.001
Pedestrian gender (Male as baseline)				
Female	-0.074	0.929	0.021	<0.001
Pedestrian age (35-44 as baseline)				
0-14	-0.391	0.676	0.032	<0.001
45-54	0.334	1.397	0.037	<0.001
55-64	0.602	1.826	0.039	<0.001
65-74	0.305	1.357	0.040	<0.001
≥75	1.000	2.718		
SD				
Pedestrian age ≥75	**0.580**	**1.786**	**0.036**	**<0.001**
Cut points				
Cut1	0.827	2.286	0.014	
Cut2	3.828	45.971	0.035	
Log likelihood null model		-48,217.27		
Log likelihood full model		-40,068.60		
R^2McFadden		0.169		

Overall, the RPOL model exhibited 64% of correct classification for slight injury crashes, 54% for serious injury, and 10% for fatal crashes with a global accuracy superior to 61%. The RPOL model, considering the most frequent class to assess the total correct classification the model global accuracy increased compared with the standard OL. However, ROL performance was lower than MNL and RPMNL models' performances.

Table 54 – Confusion matrix for the ordered logit, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	31,837	17,698	96
	Serious	7,477	8,754	128
	Fatal	150	1,080	136

Table 55 – Performance of metrics for the ordered logit model, Great Britain.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.630	0.995	0.570
TP_{rate} (Acc+)	0.535	0.100	0.605
Precision	0.318	0.378	0.679
F-measure	0.399	0.158	0.627
G-mean	0.581	0.315	0.580
AUC	0.616	0.854	0.645
Acc	0.605		
Err	0.395		

5.1.5 Classification tree

The Classification tree obtained for Great Britain is reported in Figure 19.

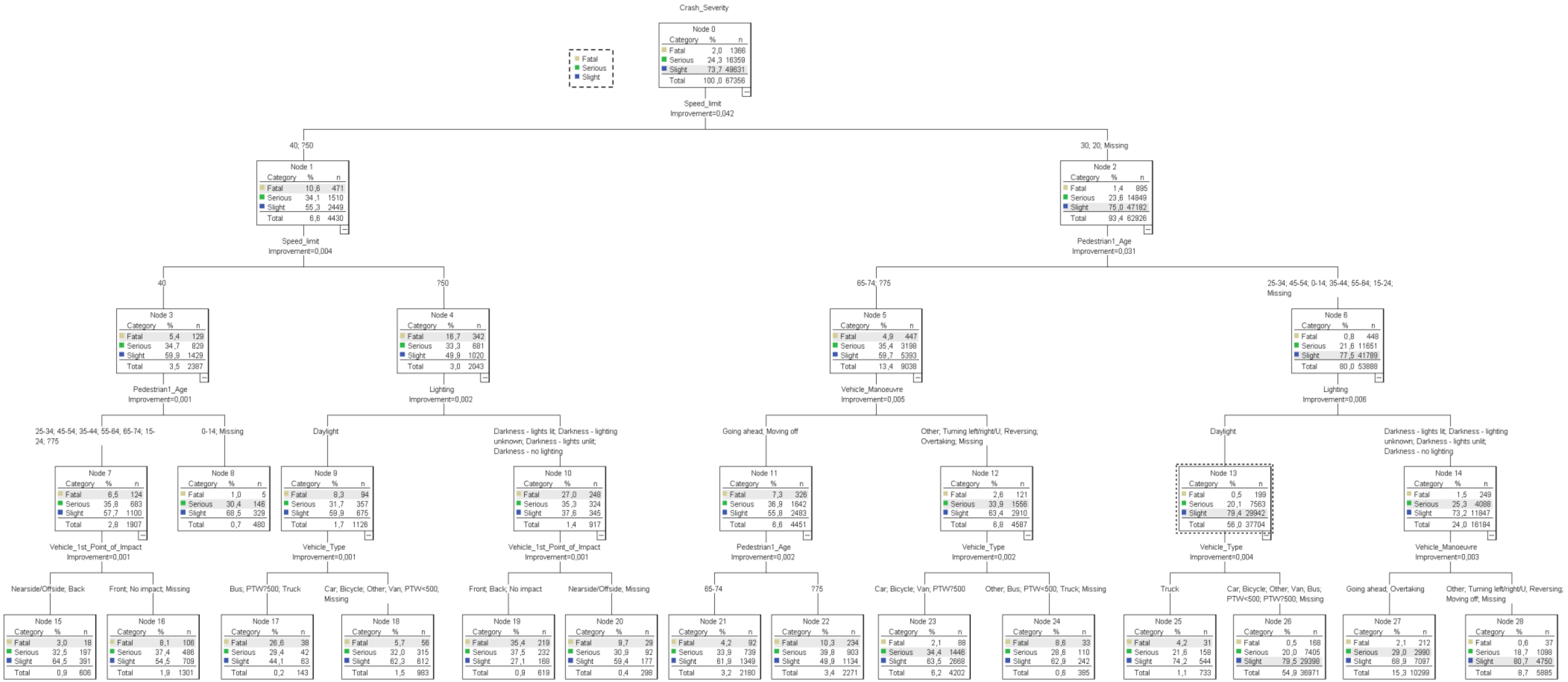


Figure 19 – Classification tree, Great Britain.

The tool generated 15 terminal nodes, 10 of which predicted fatal crashes, 3 predicted serious crashes, and 2 predicted slight injury crashes.

The posterior classification ratio (PCR) was assessed for all the nodes (see APPENDIX 1 ~ GREAT BRITAIN, classification tree section) but was reported only for the terminal nodes to understand how representative each terminal node is in relation to the predicted class. Node 17 and node 19 exhibited a very high PCR (13.10 and 17.45 respectively) which implies the robustness of both the terminal nodes for fatal classification.

Table 56 – Terminal nodes and relative Posterior Classification Ratio value, Great Britain.

	PCR			
Terminal Nodes	**Fatal**	**Serious**	**Slight**	**Actual Predicted Class**
8	0.51	1.25	0.93	Serious
15	1.46	1.34	0.88	Fatal
16	4.02	1.54	0.74	Fatal
17	13.10	1.21	0.60	Fatal
18	2.81	1.32	0.84	Fatal
19	17.45	1.54	0.37	Fatal
20	4.80	1.27	0.81	Fatal
21	2.08	1.40	0.84	Fatal
22	5.08	1.64	0.68	Fatal
23	1.03	1.42	0.86	Serious
24	4.23	1.18	0.85	Fatal
25	2.09	0.89	1.01	Fatal
26	0.22	0.82	1.08	Slight
27	1.02	1.20	0.94	Serious
28	0.31	0.77	1.10	Slight

The analysis of variable importance (Figure 20) identified four variables as mostly influencing the classification accuracy of pedestrian crash severity: (1) speed limit, (2) pedestrian age, (3) lighting, and (4) area.

Table 57 – CT Independent Variable Importance, Great Britain.

Independent Variable	Importance	Normalized Importance
Speed Limit	0.046	100.0%
Pedestrian Age	0.037	80.8%
Lighting	0.033	71.1%
Area	0.025	54.7%
Vehicle Towing and Articulation	0.016	35.5%
Road Type	0.013	28.0%
Vehicle Manoeuvre	0.010	20.8%
1st Road Class	0.009	20.5%
Vehicle Type	0.008	16.9%
Pedestrian Movement	0.004	8.9%
Vehicle 1st Point of Impact	0.004	7.8%
Junction Detail	0.003	6.8%
Pedestrian Location	0.003	6.0%
Vehicle Engine Capacity (CC)	0.003	5.5%
Pedestrian Crossing Physical Facilities	0.002	4.8%
Pavement	0.002	4.6%
X2nd Road Class	0.002	3.4%
Vehicle Junction Location	0.002	3.4%
Day of Week	0.001	3.0%
Vehicle Skidding and Overturning	0.001	2.2%
Driver Journey Purpose	0.001	1.9%
Driver Age	0.001	1.4%
Junction Control	0.001	1.3%
Pedestrian Crossing Human Control	0.001	1.2%
Weather	<0.001	0.7%
Number of Pedestrian	<0.001	0.3%
Vehicle Propulsion Code	<0.001	0.2%
Pedestrian Gender	<0.001	0.0%

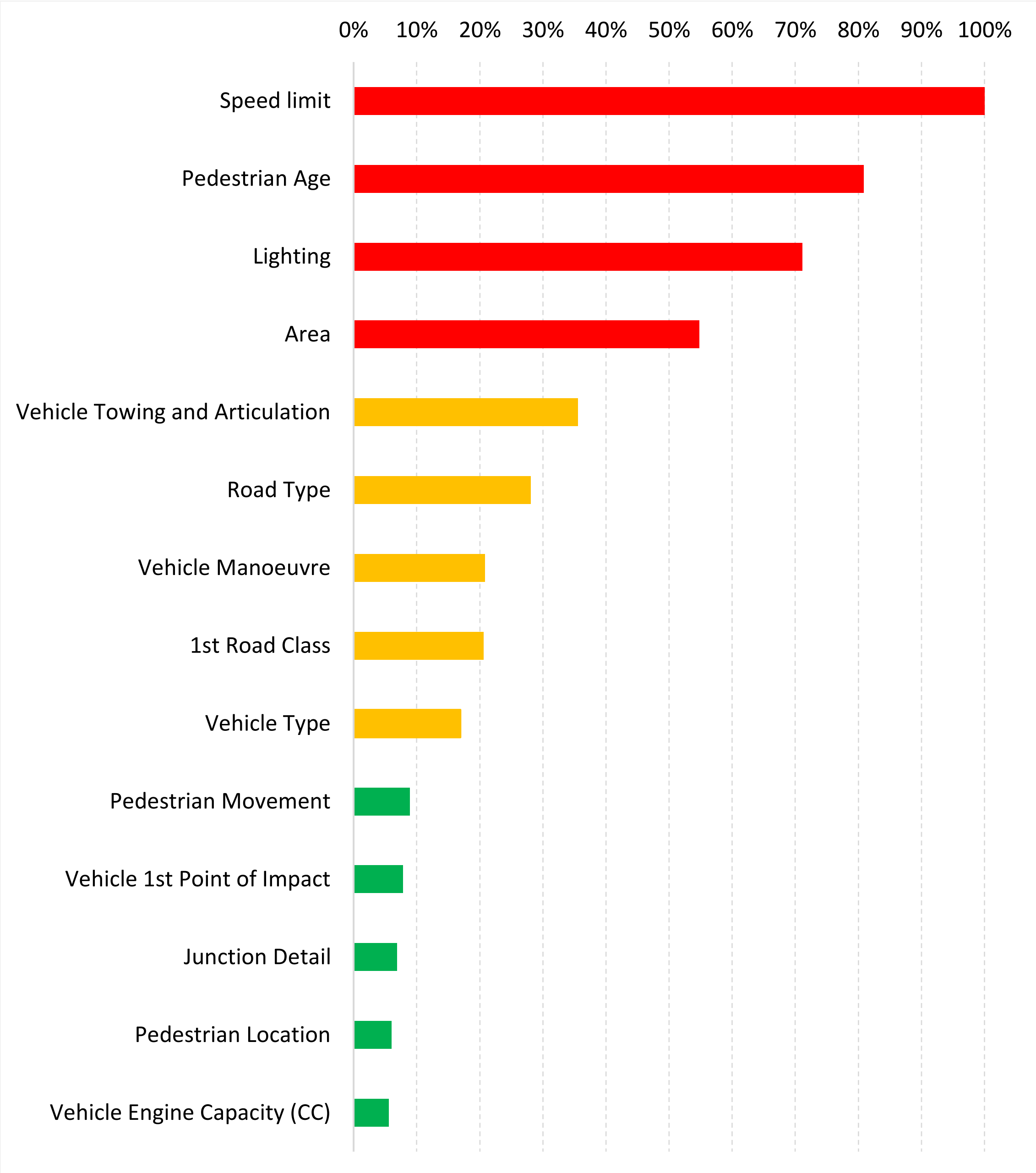


Figure 20 – CT variable normalized importance, Great Britain.

Overall, the CT tool exhibited 69% of correct classification for slight injury crashes, 28% for serious injury, and 63% for fatal crashes with a global accuracy superior to 59%.

Table 58 – Confusion matrix for the CT, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	34,148	10,094	5389
	Serious	8,503	4,582	3274
	Fatal	205	305	856

Table 59 – Performance of metrics for the CT, Great Britain.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.771	0.817	0.487
TP_{rate} (Acc+)	0.280	0.627	0.588
Precision	0.306	0.090	0.663
F-measure	0.292	0.157	0.618
G-mean	0.46	0.716	0.506
AUC	0.465	0.822	0.579
Acc	0.588		
Err	0.412		

5.1.6 Random forest

Initially, RF was implemented generating 500 trees. However, after setting the optimal number of trees based on the out of bag (OOB) sample error rate, the RF tool was performed with 42 trees. After determining the number of optimal trees, RF was performed again to determine the list of the most important variables associated with pedestrian crash severity. Indeed, the most interesting output provided by the algorithm consists in the estimates of the importance of the predictor variables for fatal and serious crashes distinctly. The importance of each explanatory variable is automatically assessed by observing the prediction error increase when the data not in the bootstrap sample (what Breiman calls OOB data) for that variable are permuted while all others are left unchanged.

All 42 trees were extracted. However, below only the first tree generated by the random forest has been reported (Figure 21). The other trees were reported in the APPENDIX 1 ~ GREAT BRITAIN (random forest section).

The random forest tool also provides the score ranking of the importance of each explanatory variable in generating the forest. Furthermore, the importance of each variable was provided for fatal and serious injury classifications separately. The normalized importance of the variables (VIMP) was reported both in table format (Table 60 and Table 61) and as figures (Figure 22 and Figure 23). According to both the Gini impurity, four variables were identified as mostly influencing the classification accuracy of fatal pedestrian crashes: vehicle manoeuvre, pedestrian age, vehicle 1st point of impact, and driver gender whereas, as far as serious crashes are concerned, RF highlighted the severe impact on pedestrian crash severity of factors such as vehicle manoeuvre and driver gender and identified as critical also the presence of vehicle towing and articulation and the vehicle type. Some variables as vehicle manoeuvre and driver gender have a significant impact on both level of crash severity exhibiting a normalized importance in generating the random forest superior to 50% (specifically equal to 100.0% and 59.3% for fatal classification, and 100.0% and 67.7% for serious injury classification respectively).

For some variables, the importance was close to zero or negative. The first case, (VIMP equal to 0) indicates the variable contributes nothing to predictive accuracy whereas the negative values indicate the predictive accuracy improves when the variable is misspecified. In the latter case, the noise is assumed being more informative than the true variable. As such, the variables with negative and near zero values of VIMP were ignored, relying on large positive values to indicate that the predictive power of the forest is dependent on those variables (Strickland, 2017).

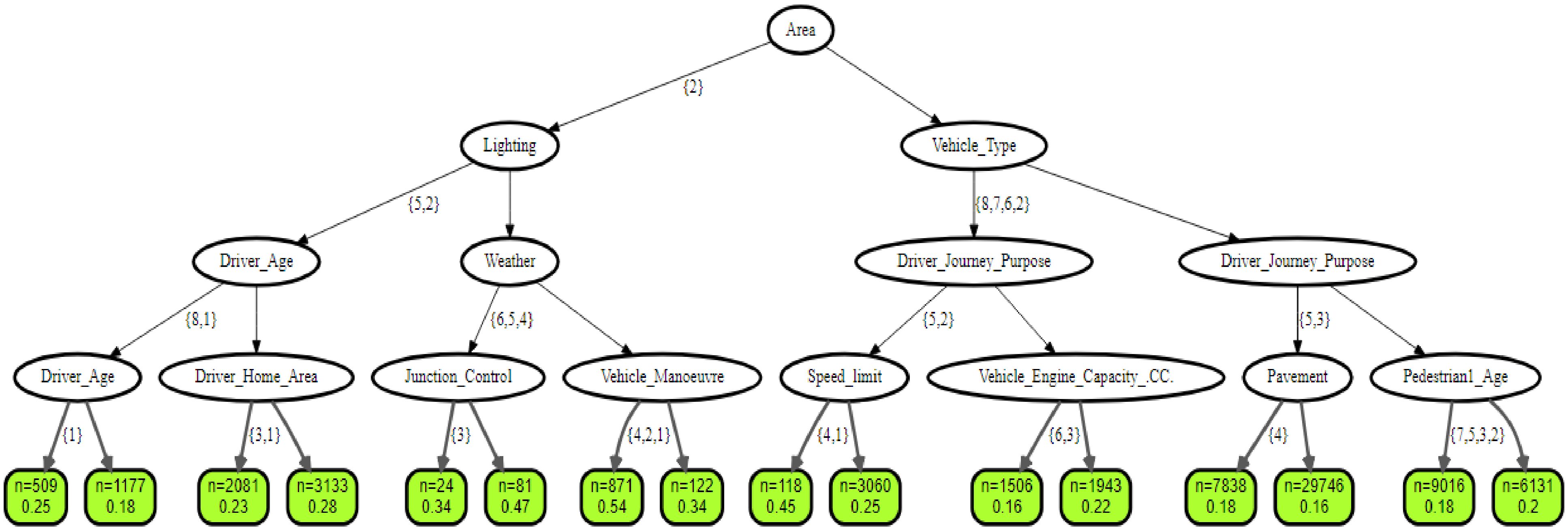


Figure 21 – First tree of the Random Forest, Great Britain.

Table 60 – RF Independent Variable Importance for fatal classification, Great Britain.

Fatal Independent Variables	Importance	Normalized importance
Vehicle Manoeuvre	0.029	100.0%
Pedestrian Age	0.022	76.5%
Vehicle 1st Point of Impact	0.019	67.0%
Driver Gender	0.017	59.3%
Driver Age	0.011	36.8%
1st Road Class	0.010	33.5%
Driver Home Area	0.007	23.8%
Weather	0.006	19.2%
Pedestrian Movement	0.005	17.2%
Junction Detail	0.003	10.5%
Pedestrian Gender	0.002	7.2%
Vehicle Propulsion Code	0.002	6.9%
Vehicle Age	0.002	6.1%
2nd Road Class	0.002	5.4%
Driver Journey Purpose	0.001	3.9%
Pedestrian Crossing Human Control	0.001	3.0%
Pedestrian IMD Decile	0.001	2.8%
Vehicle Junction Location	0.001	2.0%
Driver IMD Decile	0.000	-0.5%
Road Type	0.000	-1.0%
Junction Control	0.000	-1.1%
Day of Week	0.000	-1.5%
Pedestrian Crossing Physical Facilities	0.000	-1.6%
Pedestrian Location	-0.001	-2.0%
Pavement	-0.001	-2.1%
Area	-0.001	-3.1%
Vehicle Engine Capacity (CC)	-0.001	-4.1%
Number of Vehicles	-0.002	-7.1%
Number of Pedestrian	-0.003	-10.5%
Vehicle Skidding and Overturning	-0.012	-40.1%
Vehicle Type	-0.019	-66.0%
Lighting	-0.021	-71.6%
Speed Limit	-0.021	-73.7%
Vehicle Towing and Articulation	-0.030	-105.4%

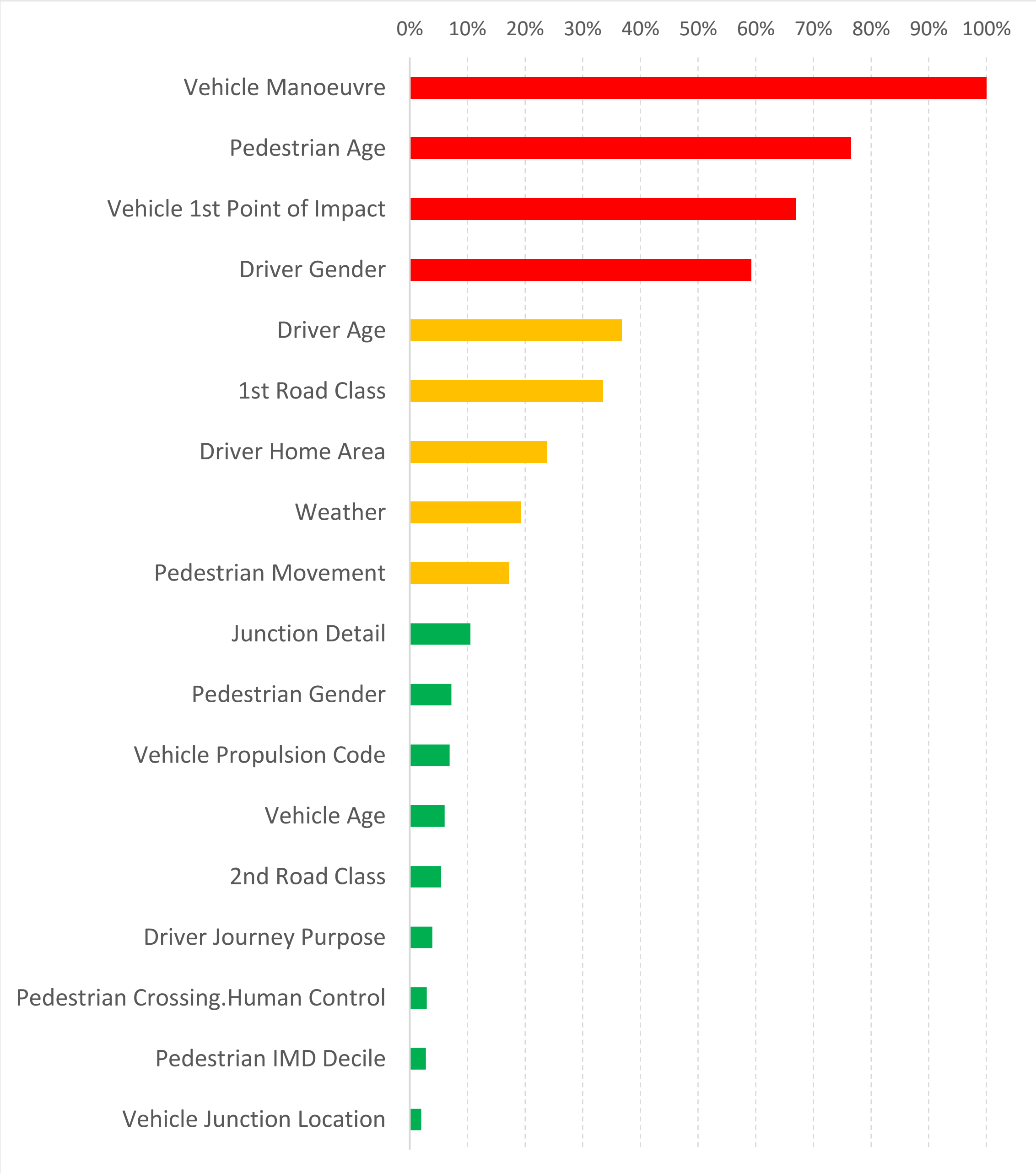


Figure 22 – RF variable normalized importance for fatal classification, Great Britain.

Table 61 – RF Independent Variable Importance for serious injury classification, Great Britain.

Serious Independent Variables	Importance	Normalized importance
Vehicle Manoeuvre	0.076	100.0%
Vehicle Towing and Articulation	0.069	91.3%
Driver Gender	0.051	67.7%
Vehicle Type	0.039	51.7%
Speed limit	0.035	46.7%
Lighting	0.032	42.7%
Vehicle 1st Point of Impact	0.032	42.1%
Driver Home Area	0.025	32.7%
Driver Age	0.024	31.1%
Driver IMD Decile	0.017	23.0%
Pedestrian Location	0.012	15.8%
Area	0.011	14.6%
Road Type	0.011	13.9%
Vehicle Propulsion Code	0.009	11.7%
1st Road Class	0.004	5.1%
Pedestrian Crossing Human Control	0.003	4.2%
Vehicle Age	0.003	4.1%
Junction Detail	0.001	1.7%
Junction Control	0.001	1.1%
Pedestrian Gender	0.000	0.3%
Pedestrian IMD Decile	0.000	-0.5%
Vehicle Junction Location	0.000	-0.7%
2nd Road Class	-0.001	-0.9%
Day of Week	-0.001	-1.3%
Pavement	-0.003	-3.9%
Vehicle Engine Capacity (CC)	-0.006	-7.6%
Pedestrian Movement	-0.006	-7.8%
Number of Pedestrian	-0.006	-8.6%
Pedestrian Crossing Physical Facilities	-0.008	-10.1%
Vehicle Skidding and Overturning	-0.009	-11.6%
Weather	-0.010	-12.7%
Number of Vehicles	-0.020	-26.6%
Pedestrian Age	-0.028	-36.6%
Driver Journey Purpose	-0.029	-38.3%

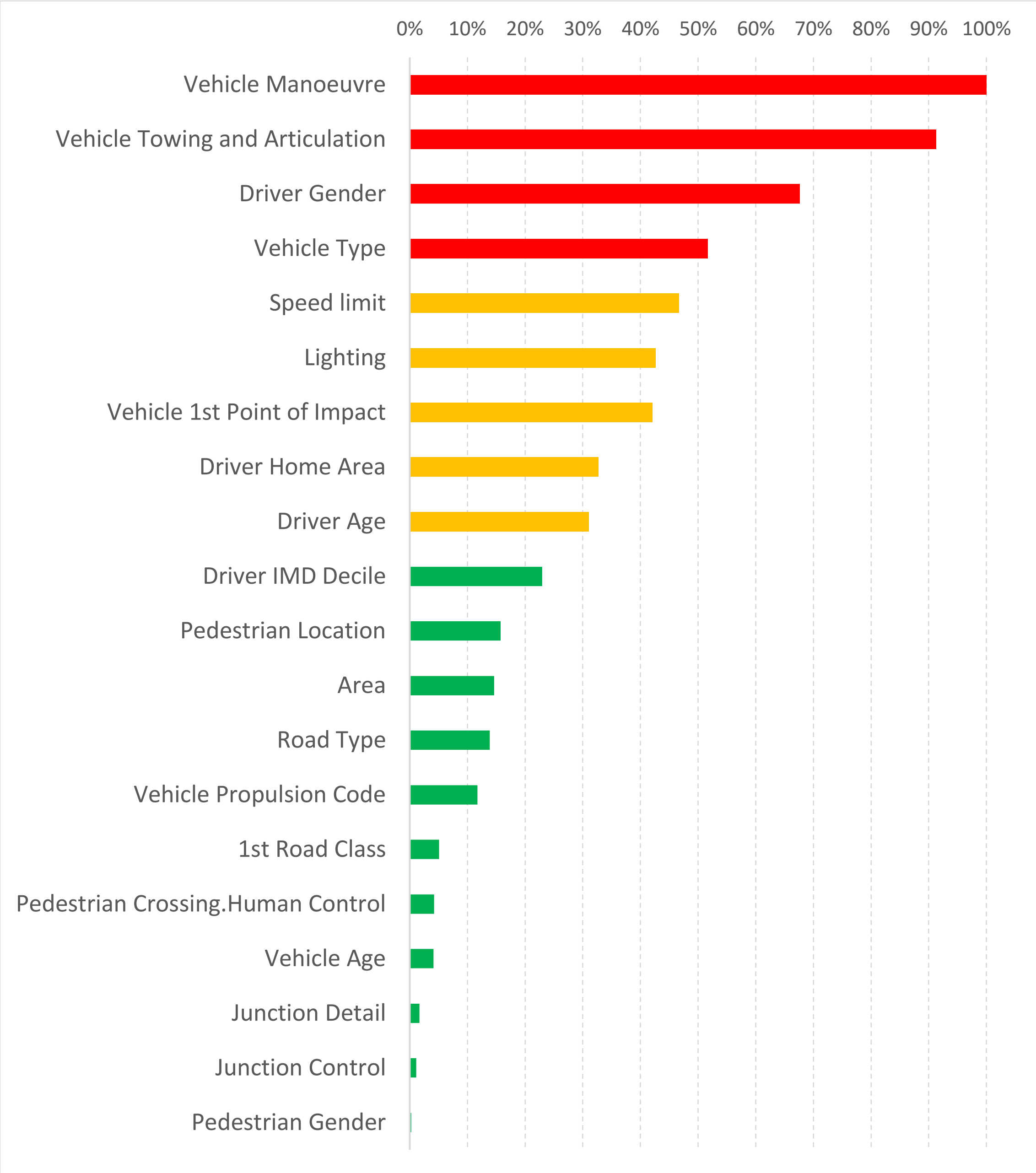


Figure 23 – RF variable normalized importance for serious injury classification, Great Britain.

Overall, the RF tool exhibited 76% of correct classification for slight injury crashes, 36% for serious injury, and 41% for fatal crashes with a global accuracy superior to 66%. As expected, the global accuracy of RF tool was superior, even slightly, to the global accuracy exhibited by CT algorithm.

Table 62 – Confusion matrix for the RF, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	37,804	10,464	1,363
	Serious	9,437	5,817	1,105
	Fatal	238	575	553

Table 63 – Performance metrics for RF, Great Britain.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^{-})$	0.777	0.946	0.517
TP_{rate} (Acc^{+})	0.356	0.405	0.644
Precision	0.345	0.183	0.656
Recall	0.356	0.405	0.644
F-measure	0.350	0.252	0.648
G-mean	0.525	0.619	0.547
AUC	0.532	0.866	0.627
Acc	0.656		
Err	0.344		

5.1.7 Association rules

The threshold values of support (S), confidence (C), and lift (L) were set as follows: S ≥ 0.1%, C ≥ 4%, L ≥ 1.2, and LIC ≥ 1.05.

The a priori algorithm generated 254 rules with fatal crash as consequent and 475 rules with serious crash as consequent. Furthermore, the extracted rules exhibited at most three items as antecedents. Among the rules with fatal crash as consequent, 97 rules included pedestrian age not smaller than 75 as first antecedent, 53 rules included vehicle engine capacity not smaller than 3000, 33 rules included rural area, 26 included vehicle skidding and overturning, and 15 included lighting equal to darkness - no lighting. The rules were organized in such a way that the 2-item rules were ordered by the decreasing value of lift, the 3-item rules having the same antecedent of the 2-item rule were ordered again by the decreasing value of the lift, and so on and then the rules were grouped according to the strongest 2-item parent rules.

Pedestrian age generated a considerable number of significant rules also for serious injury as consequent. Out of the 475 rules with serious crash as consequent, 237 rules exhibited pedestrian age as the first item, followed by 74 rules with the number of pedestrians involved in a crash and drivers aged under 25 with 33 rules. Table 64 and Table 65 contain the strongest rules predicting fatal and serious crashes.

Table 64 – Association rules with fatal as consequent, Great Britain.

Antecedents	**S %**	**C %**	**Lift**	**LIC**
Vehicle towing and articulation=Yes	0.14	28.87	14.24	n.a.
Lighting=Darkness - no lighting	0.33	17.80	8.78	n.a.
Lighting=Darkness - no lighting & Speed limit ≥ 50	0.29	30.06	14.82	1.69
Speed limit ≥ 50	0.51	16.74	8.25	n.a.
Speed limit ≥ 50 & Day of week=Weekend	0.16	18.41	9.08	1.10
Vehicle type=Truck	0.30	13.64	6.73	n.a.
Vehicle skidding and overturning=Yes	0.21	7.63	3.76	n.a.
Pedestrian age ≥ 75	0.56	7.46	3.68	n.a.
Pedestrian age ≥ 75 & Lighting=Darkness - lights lit	0.15	13.96	6.88	1.87
Pedestrian age ≥ 75 & Lighting=Darkness - lights lit & Vehicle 1st point of impact=Front	0.12	16.94	8.35	1.21
Pedestrian age ≥ 75 & Lighting=Darkness - lights lit & Driver home area=Urban	0.11	14.72	7.26	1.05
Pedestrian age ≥ 75 & Lighting=Darkness - lights lit & Vehicle age ≥ 15	0.11	14.68	7.24	1.05
Pedestrian age ≥ 75 & Vehicle Manoeuvre=Going ahead	0.37	12.30	6.07	1.65
Pedestrian age ≥ 75 & Vehicle Manoeuvre=Going ahead & Pavement=Wet or damp	0.11	14.14	6.97	1.15
Pedestrian age ≥ 75 & Vehicle Manoeuvre=Going ahead & Vehicle propulsion =Petrol	0.18	13.87	6.84	1.13
Pedestrian age ≥ 75 & Vehicle Manoeuvre=Going ahead & Junction detail=T or staggered	0.12	13.42	6.62	1.09
Pedestrian age ≥ 75 & Vehicle 1st Point of Impact=Front	0.40	10.41	5.13	1.40
Pedestrian age ≥ 75 & Vehicle 1st Point of Impact=Front & Junction Control=Not at junction or within 20 metres	0.18	13.16	6.49	1.26
Pedestrian age ≥ 75 & Vehicle 1st point of impact=Front & Vehicle Propulsion=Heavy oil	0.17	12.71	6.27	1.22
Pedestrian age ≥ 75 & Vehicle 1st point of impact=Front & Vehicle age ≥ 15	0.30	11.18	5.51	1.07
Pedestrian age ≥ 75 & Day of week=Weekend	0.14	9.76	4.81	1.31
Pedestrian age ≥ 75 & Day of week=Weekend & Driver gender=M	0.10	11.09	5.47	1.14
Pedestrian age ≥ 75 & Driver journey purpose=Journey as part of work	0.16	9.70	4.79	1.30
Pedestrian age ≥ 75 & Pavement=Wet or damp	0.15	8.88	4.38	1.19
Pedestrian age ≥ 75 & Vehicle Propulsion=Heavy oil	0.25	8.82	4.35	1.18
Pedestrian age ≥ 75 & Driver gender=M	0.43	8.74	4.31	1.17
Pedestrian age ≥ 75 & Pedestrian gender=M	0.31	8.47	4.17	1.13
Pedestrian age ≥ 75 & Driver age=25-34	0.11	8.10	3.99	1.09
Vehicle engine capacity ≥3000	0.35	6.89	3.40	n.a.
Vehicle engine capacity ≥3000 & Speed limit ≥ 50	0.10	39.53	19.49	5.74
Vehicle engine capacity ≥3000 & Driver journey purpose=Journey as part of work	0.31	8.17	4.03	1.19
Vehicle engine capacity ≥3000 & Driver gender=M	0.33	7.33	3.61	1.06
Area=Rural	0.68	5.71	2.82	n.a.
Area=Rural & Number of vehicles=2	0.10	10.15	5.00	1.78
Area=Rural & Day of week=Weekend	0.22	8.04	3.96	1.41

Table 65 – Association rules with serious crashes as consequent, Great Britain.

Antecedents	S %	C %	Lift	LIC
Number of Pedestrians involved ≥ 2	0.14	42.48	1.75	n.a.
Pedestrian age ≥ 75	2.82	37.35	1.54	n.a.
Pedestrian age ≥ 75 & Vehicle age ≥ 1	0.18	46.88	1.93	1.26
Pedestrian age ≥ 75 & Driver journey purpose=Commuting to/from work	0.26	44.53	1.83	1.19
Pedestrian age ≥ 75 & Pavement=Wet or damp	0.74	42.93	1.77	1.15
Pedestrian age ≥ 75 & Driver age ≥ 75	0.29	42.49	1.75	1.14
Pedestrian age ≥ 75 & Driver home area=Small town	0.22	42.3	1.74	1.13
Pedestrian age ≥ 75 & Pedestrian crossing physical facilities=Zebra	0.2	41.77	1.72	1.12
Pedestrian age ≥ 75 & Pedestrian crossing physical facilities =Zebra & Driver gender=M	0.15	46.7	1.92	1.12
Pedestrian age ≥ 75 & Vehicle type=Van	0.27	40.77	1.68	1.09
Pedestrian age ≥ 75 & Vehicle type=Van & Junction detail=T or staggered	0.11	48.1	1.98	1.18
Pedestrian age ≥ 75 & Vehicle type=Van & Junction control=Give way/uncontrolled	0.15	45.02	1.85	1.1
Pedestrian age ≥ 75 & Vehicle propulsion code=Petrol	1.23	40.68	1.67	1.09
Pedestrian age ≥ 75 & Pedestrian gender=F	1.58	40.54	1.67	1.09
Vehicle Skidding and Overturning=Yes	0.97	35.37	1.46	n.a.
Speed limit=40	1.23	34.73	1.43	n.a.
Speed limit=40 & Day of week=Weekend	0.32	39.63	1.63	1.14
Pedestrian age=65-74	2.22	33.41	1.38	n.a.
Pedestrian age=65-74 & Driver journey purpose=Commuting to/from work	0.21	42.22	1.74	1.26
Pedestrian age=65-74 & Driver age=0-24	0.27	39.57	1.63	1.18
Pedestrian age=65-74 & Driver age=0-24 & Vehicle age ≥ 15	0.22	42.44	1.75	1.07
Pedestrian age=65-74 & Pavement=Wet or damp	0.63	37.63	1.55	1.13
Lighting=Darkness - no lighting	0.61	33.2	1.37	n.a.
Lighting=Darkness - no lighting & Speed limit ≥ 50	0.34	35.51	1.46	1.07
Weather=Raining + high winds	0.31	31.09	1.28	n.a.
Driver age=0-24	3.06	29.32	1.21	n.a.
Driver age=0-24 & Speed limit ≥ 50	0.14	38.56	1.59	1.31
Driver age=0-24 & Speed limit ≥ 50 & Vehicle 1st point of impact=Front	0.1	41.72	1.72	1.08
Driver age=0-24 & Day of week=Weekend	0.81	31.21	1.29	1.06
Lighting=Darkness - lights unlit	0.22	29.32	1.21	n.a.

Overall, the AR tool exhibited 51% of correct classification for slight injury crashes, 52% for serious injury, and 82% for fatal crashes with a global accuracy superior to 51%. Even though the global accuracy reached a very low value, the accuracy in correctly classification of serious injury and fatal crashes considerably increase. The improvement is significant especially for fatal with 82% cases correctly classified as fatal.

Table 66 – Confusion matrix for the RF, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	25,344	18,438	5,849
	Serious	6,117	8,533	1,709
	Fatal	159	990	217

Table 67 – Performance metrics for RF, Great Britain.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^{-})$	0.582	0.568	0.818
TP_{rate} (Acc^{+})	0.511	0.522	0.159
Precision	0.802	0.305	0.028
Recall	0.624	0.385	0.047
F-measure	0.545	0.544	0.360
G-mean	0.610	0.578	0.790
AUC	0.582	0.568	0.818
Acc	0.506		
Err	0.494		

5.1.8 Support vector machine

SVM model was performed with RBF kernel function. The model returned 19,909 support vectors defining the complex hyperplane. Among the output of the tool, SVM provides the visualization of the most relevant features through non-linear kernels necessary to carry out the classification process. To compare SVM output with the outputs of the other machine learning algorithms implemented in the study, we reported this visualization of the most important predictors exhibited by the tool. SVM identified 4 predictors mostly contributing to the correct classification of pedestrian crash severity: 1st road class, pedestrian age, pedestrian crossing physical facilities, and junction detail.

Table 68 – SVM Independent Variable Importance, Great Britain.

Independent Variables	Importance	Normalized Importance
1st Road Class	1.000	100.0%
Pedestrian Age	0.616	61.6%
Pedestrian Crossing Physical Facilities	0.615	61.5%
Junction Detail	0.586	58.6%
Vehicle Towing and Articulation	0.438	43.8%
Vehicle Type	0.400	40.0%
Number of Vehicles	0.392	39.2%
Speed Limit	0.311	31.1%
Driver Age	0.273	27.3%
Vehicle Skidding and Overturning	0.258	25.8%
Pedestrian Crossing Human Control	0.207	20.7%
Pavement	0.188	18.8%
Vehicle Manoeuvre	0.155	15.5%
Vehicle Propulsion Code	0.083	8.3%
Lighting	0.048	4.8%
Vehicle Age	0.039	3.9%
Area	0.036	3.6%
Vehicle Junction Location	0.023	2.3%
Pedestrian Gender	0.023	2.3%
Driver Gender	0.017	1.7%
Day of Week	0.017	1.7%

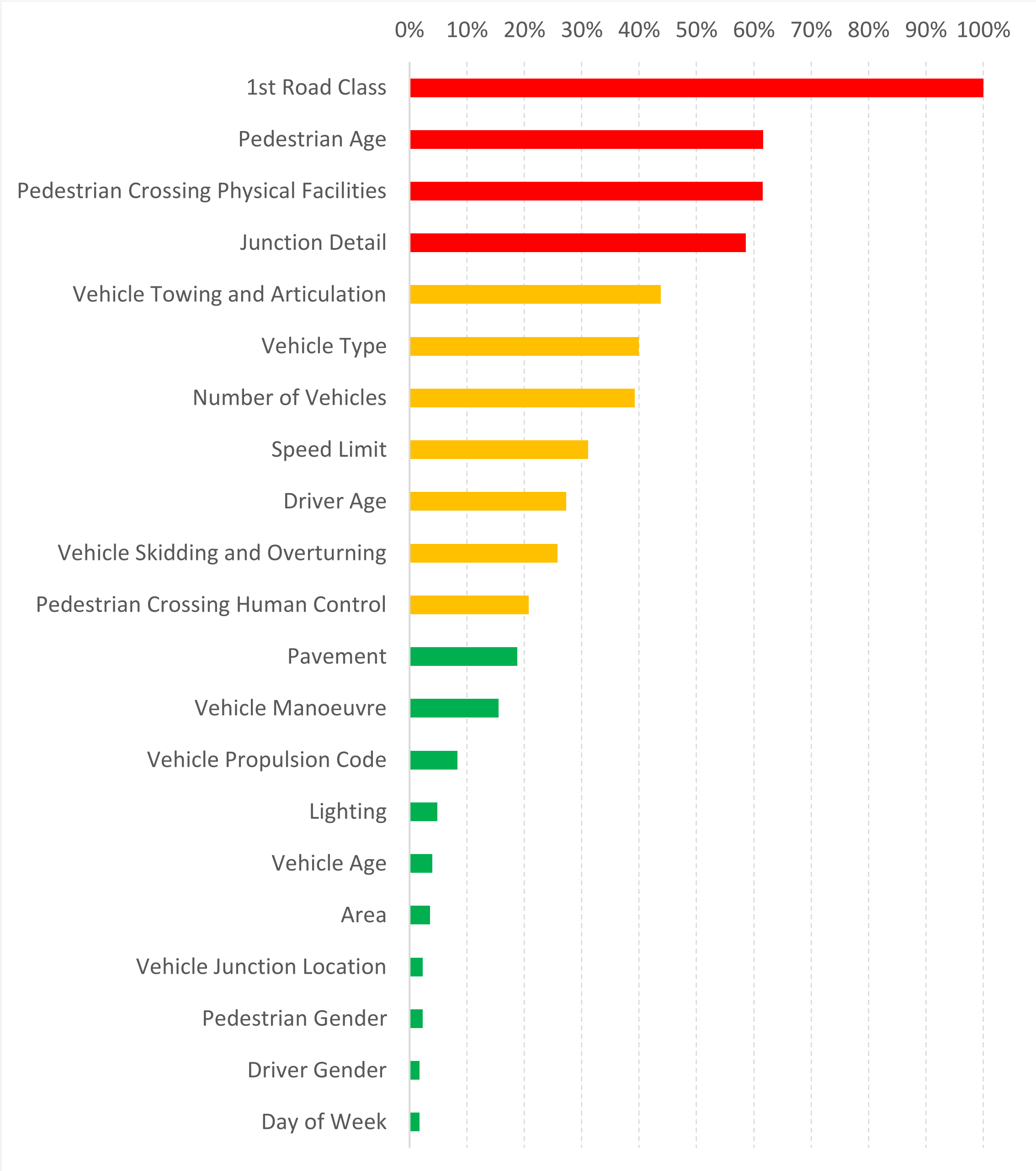


Figure 24 – SVM variable importance, Great Britain.

Overall, the SVM tool exhibited 93% of correct classification for slight injury crashes, 100% for serious injury, and 100% for fatal crashes with a global accuracy equal to 98%. SVM reached the highest classification accuracy.

Table 69 – Confusion matrix for SVM, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	10,092	40	0
	Serious	285	3,405	0
	Fatal	16	11	282

Table 70 – Performance metrics for SVM, Great Britain.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^{-})$	0.995	1.000	0.943
TP_{rate} (Acc^{+})	0.923	0.913	0.977
Precision	0.985	1.000	0.975
F-measure	0.953	0.954	0.975
G-mean	0.958	0.955	0.959
AUC	0.758	0.879	0.772
Acc	0.975		
Err	0.025		

5.1.9 Artificial neural network

ANN generated a graph containing 26 factors and 132 neurons in the input layer (excluding the bias unit) whereas the output layer had three neurons that represent the 3 injury levels. All transfer functions at the input-hidden layers were hyperbolic tangent transfer functions whereas the transfer function was the softmax function between the hidden layer and the output layer. 13 factors exhibited high impact on pedestrian crash severity with a normalized importance greater than 50%: driver and pedestrian age, vehicle engine, lighting, vehicle 1st point of impact, speed limit, vehicle manoeuvre, vehicle type, area, 1st road class, weather, junction detail, and pedestrian crossing physical facilities.

The parameter estimates provided by the tool were reported in the appendix (the reader refers to APPENDIX 1 ~ GREAT BRITAIN, Artificial Neural Network section).

Table 71 – Artificial Neural Network general information, Great Britain.

Network Information			
Input Layer	Factors	1	Number of Vehicles
		2	Day of Week
		3	1st Road Class
		4	Road Type
		5	Speed limit
		6	Junction Detail
		7	Junction Control
		8	Pedestrian Crossing Human Control
		9	Pedestrian Crossing Physical Facilities
		10	Lighting
		11	Weather
		12	Pavement
		13	Area
		14	Vehicle Type
		15	Vehicle Towing and Articulation
		16	Vehicle Manoeuvre
		17	Vehicle Junction Location
		18	Vehicle Skidding and Overturning
		19	Vehicle 1st Point of Impact
		20	Driver Gender
		21	Driver Age
		22	Vehicle Engine Capacity
		23	Vehicle Propulsion Code
		24	Vehicle Age
		25	Pedestrian Gender
		26	Pedestrian Age
	Number of Units[a]		132
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		13
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Crash Severity
	Number of Units		3
	Activation		Softmax
	Function Error Function		Cross-entropy

a. Excluding the bias unit

Table 72 – ANN Independent Variable Importance, Great Britain.

Independent variables	Importance	Normalized Importance
Driver Age	0.071	100.0%
Pedestrian Age	0.063	88.6%
Vehicle Engine Capacity (CC)	0.058	81.3%
Lighting	0.054	76.7%
Vehicle 1st Point of Impact	0.051	72.3%
Speed limit	0.051	71.5%
Vehicle Manoeuvre	0.050	70.5%
Vehicle Type	0.048	68.3%
Area	0.048	67.6%
1st Road Class	0.044	61.7%
Weather	0.043	61.1%
Junction Detail	0.040	57.2%
Pedestrian Crossing Physical Facilities	0.037	52.9%
Pavement	0.032	45.6%
Road Type	0.031	44.4%
Vehicle Age	0.031	44.3%
Vehicle Skidding and Overturning	0.031	44.1%
Driver Gender	0.030	42.2%
Vehicle Junction Location	0.027	38.7%
Number of Vehicles	0.027	38.2%
Junction Control	0.027	37.7%
Vehicle Propulsion Code	0.026	36.6%
Vehicle Towing and Articulation	0.025	34.9%
Pedestrian Crossing Human	0.022	30.9%
Control Pedestrian Gender	0.019	27.0%
Day of Week	0.013	18.3%

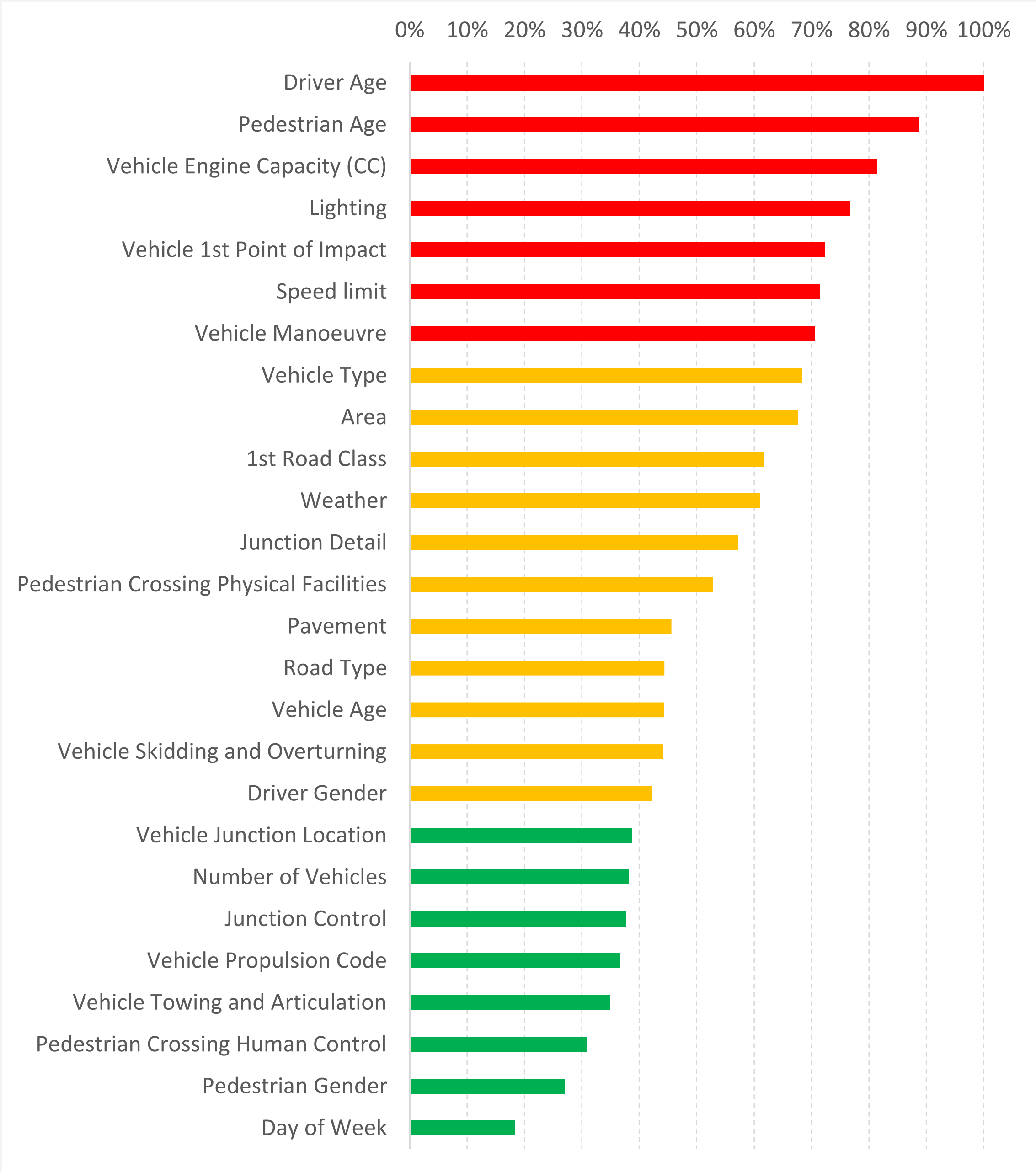


Figure 25 – ANN variable importance, Great Britain.

Overall, the ANN tool exhibited 74% of correct classification for slight injury crashes, 23% for serious injury, and 50% for fatal crashes with a global accuracy equal to 62%.

Table 73 – Confusion matrix for ANN, Great Britain.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	36,804	9,396	3,431
	Serious	10,391	3,828	2,140
	Fatal	385	295	686

Table 74 – Performance metrics for ANN, Great Britain.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^-)$	0.795	0.879	0.450
TP_{rate} (Acc^+)	0.234	0.502	0.617
Precision	0.283	0.110	0.613
F-measure	0.256	0.180	0.603
G-mean	0.431	0.665	0.479
AUC	0.761	0.782	0.810
Acc	0.613		
Err	0.387		

5.1.10 Synthesis of the results

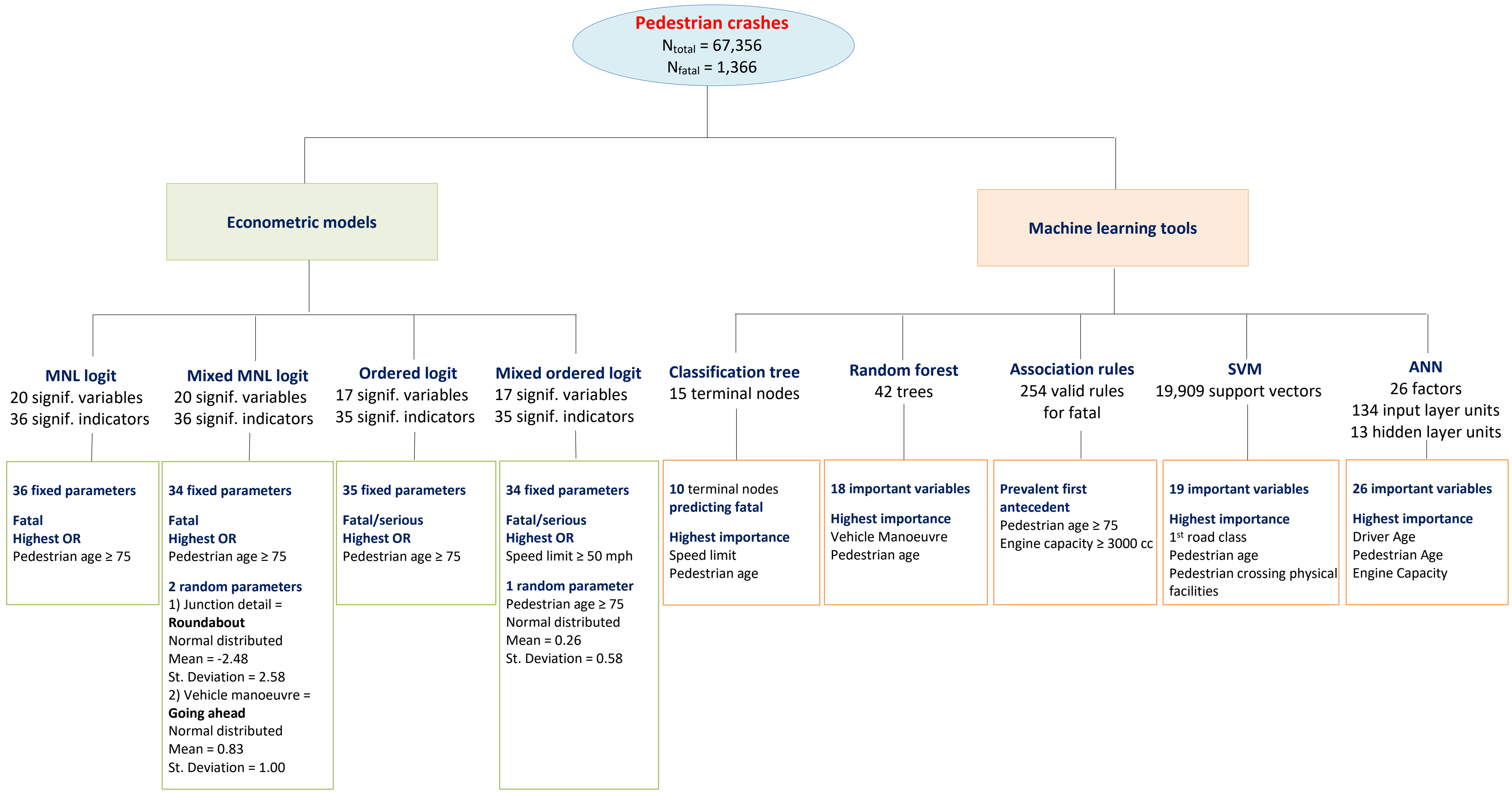


Figure 26 – Main results of the econometric models and machine learning tools for fatal crashes, Great Britain.

5.1.11 Measures of performance

For each model, TN_{rate}, TP_{rate}, precision, F-measure, G-mean, AUC, accuracy, and error rate were evaluated and reported in their result section respectively. However, in light of the considerations expressed in paragraph 3.4 (Measures of performance), in this research, the models' performances were compared by F-measure, G-mean, and AUC. The three metrics, indeed, combine different simple metrics into a single measure providing a more comprehensive framework on the performance exhibited by the models and overcoming the issue related to the accuracy (a globally model indicator of performance) and the error rate (which is the one's complement representation of accuracy).

Results are shown in Table 75-77. Table 75 reports the performance measures exhibited by econometric models both in their standard formulation, without applying any treatment of imbalanced data, as well as in their weighted formulation, after the implementation of the weighted approach presented in paragraph 3.3 of this thesis. Table 76 reports the performance measures of the machine learning algorithms in the standard and weighted formulation. After the implementation of the weighted approach, all the methods exhibited a relevant improvement in the classification performances except for ARs where the weighted formulation did not significantly affect the model's performances (Table 77). However, in a certain way, AR already allows the analyst to assign a greater "weight" to fatal and/or serious injury levels by setting initial parameters of support, confidence, lift, and lift increase.

Table 75 – Measures of performance of standard and weighted econometric models, Great Britain.

	Standard econometric models				**Weighted econometric models**			
	MNL	RPMNL	OL	RPOL	MNL	RPMNL	OL	RPOL
Fatal								
F-measure	0.16	0.23	0.00	0.02	0.28	0.53	0.00	0.16
G-mean	0.32	0.38	0.04	0.10	0.50	0.65	0.04	0.33
AUC	0.86	0.87	0.85	0.86	0.87	0.94	0.85	0.85
Serious								
F-measure	0.06	0.32	0.05	0.14	0.21	0.41	0.41	0.40
G-mean	0.17	0.46	0.17	0.28	0.36	0.58	0.43	0.58
AUC	0.62	0.63	0.61	0.63	0.62	0.68	0.61	0.62
Averaged performances								
F-measure	0.06	0.31	0.05	0.13	0.22	0.42	0.38	0.38
G-mean	0.18	0.45	0.16	0.27	0.37	0.59	0.40	0.56
AUC	0.64	0.65	0.63	0.64	0.64	0.70	0.63	0.63

Table 76 – Measures of performance of standard and weighted machine learning algorithms, Great Britain.

	Standard ML algorithms					Weighted ML algorithms				
	AR	CT	RF	ANN	SVM	AR	CT	RF	ANN	SVM
Fatal										
F-measure	0.05	0.00	0.02	0.04	0.01	0.05	0.16	0.57	0.18	0.95
G-mean	0.36	0.00	0.09	0.15	0.07	0.36	0.72	0.77	0.66	0.96
AUC	0.79	0.80	0.23	0.83	0.76	0.79	0.82	0.88	0.78	0.88
Serious										
F-measure	0.39	0.11	0.00	0.13	0.03	0.39	0.29	0.90	0.26	0.95
G-mean	0.54	0.24	0.04	0.27	0.12	0.54	0.46	0.92	0.43	0.96
AUC	0.58	0.61	0.56	0.61	0.55	0.58	0.47	0.71	0.76	0.76
Averaged performances										
F-measure	0.36	0.10	0.00	0.12	0.02	0.36	0.28	0.87	0.25	0.95
G-mean	0.53	0.22	0.05	0.26	0.11	0.53	0.48	0.91	0.45	0.96
AUC	0.59	0.63	0.53	0.63	0.56	0.59	0.49	0.72	0.76	0.77

**note that the performances of AR are the same in standard and weighted formulations.*

Table 77 – Measures of performance of weighted econometric models and machine learning algorithms, Great Britain.

	Econometric models				Machine learning				
	MNL	RPMNL	OL	RPOL	AR	CT	RF	ANN	SVM
Fatal									
F-measure	0.28	0.53	0.00	0.16	0.05	0.16	0.57	0.18	0.95
G-mean	0.50	0.65	0.04	0.32	0.36	0.72	0.77	0.66	0.96
AUC	0.87	0.94	0.85	0.85	0.79	0.82	0.88	0.78	0.88
Serious									
F-measure	0.21	0.41	0.41	0.40	0.39	0.29	0.90	0.26	0.95
G-mean	0.36	0.58	0.43	0.58	0.54	0.46	0.92	0.43	0.96
AUC	0.62	0.68	0.61	0.62	0.58	0.47	0.71	0.76	0.76
Averaged performances									
F-measure	0.22	0.42	0.38	0.38	0.36	0.28	0.87	0.25	0.95
G-mean	0.37	0.59	0.40	0.56	0.53	0.48	0.91	0.45	0.96
AUC	0.64	0.70	0.63	0.63	0.59	0.49	0.72	0.76	0.77

The comparison among the different models shows several interesting results.

As far as the econometric models are concerned, MNL (fixed parameters) and RPMNL (mixed parameters) models exhibited better classification performances compared with their ordered version (OL and RPOL). Furthermore, the OL model showed poor ability in classifying correctly fatal crashes even after the weighting procedure. The results of the methods applied on the British database find out that, among all the econometric models implemented, RPMNL has the best predictive performances and provides additional insights on the distribution of parameters in the empirical analysis.

As far as the machine learning tools is concerned, SVM outperformed the other algorithms reaching accuracy in both correct positive and negative case classification equal to 96%. RF exhibited performances only slightly worse than SVM. ANN, AR and CT exhibited similar performances with better performance of CT in predicting fatal crashes and better performance of AR in predicting severe injury crashes.

Overall, machine learning algorithms outperformed econometric models and the best performances were reached by SVM and RF.

5.1.12 Significant explanatory variables and effects on crash severity

In Table 78 and Table 79, the significant explanatory variables associated with an increase in crash severity were summarized. Table 78 contains variables associated with an increase in fatal crash probability while Table 79 contains variables associated with an increase in serious crash probability. 19 variables are significant both in the econometric models as well as in the machine learning algorithms, 1 variable is significant only in the econometric models and 7 variables are significant only in the machine learning algorithms. This means that data-driven methods tend to uncover more hidden correlations among data than econometric models. The same variables are significant with reference to both fatal as well as serious injuries except for the vehicle propulsion code (significant only for fatal severity) and the number of pedestrians involved (significant only for serious injuries). The variable significant only in the econometric model is the pedestrian crossing human control while the variables significant only in the machine learning algorithms are driver home area, driver journey purpose, number of pedestrians involved, vehicle first point of impact, vehicle engine capacity, weather, and junction control.

Table 78 – Variables associated with an increase in fatal crash probability, Great Britain.

Econometric and ML models	Only econometric models	Only ML algorithms
1st road class	Pedestrian crossing human control	Driver home area
Area		Driver journey purpose
Day of week		Vehicle 1st point of impact
Driver age		Vehicle engine capacity
Driver gender		Weather
Lighting		Junction control
N. of vehicles involved		
Pavement		
Pedestrian age		
Pedestrian crossing physical facilities		
Pedestrian gender		
Speed limit		
Vehicle age		
Vehicle manoeuvre		
Vehicle propulsion code		
Vehicle skidding and overturning		
Vehicle towing and articulation		
Vehicle type		
Junction detail		

Table 79 – Variables associated with an increase in serious injury crash probability, Great Britain.

Econometric and ML models	Only econometric models	Only ML algorithms
1st road class	Pedestrian crossing human control	Driver home area
Area		Driver journey purpose
Day of week		N. of pedestrians involved
Driver age		Vehicle 1st point of impact
Driver gender		Vehicle engine capacity
Lighting		Weather
N. of vehicles involved		Junction control
Pavement		
Pedestrian age		
Pedestrian crossing physical facilities		
Pedestrian gender		
Speed limit		
Vehicle age		
Vehicle manoeuvre		
Vehicle skidding and overturning		
Vehicle towing and articulation		
Vehicle type		
Junction detail		

5.1.11.1 Roadway characteristics

Econometric models identified the increase of the speed limit as a contributory factor of the increase in the crash severity. The speed limit was also the first split for CT growth with higher speed limits associated with fatal crashes. AR identified high-lift rules with fatal severity as consequent and speed limit ≥ 50 mph as antecedent. The speed limit was also identified as one of the most important predictors by ANN with 70% importance. All the models also pointed out 1st road class equal to A and rural area as patterns influencing crash severity and this may be due to the correlation with higher speed limits.

Pelican, puffin, toucan or similar non-junction pedestrian light crossing were found increasing pedestrian crash severity. As far junction detail is concerned, econometric models did not provide factors influencing severity levels. By contrast, AR found T or staggered junction or give-way/uncontrolled intersection affecting fatal and serious crashes in presence of elderly pedestrians and van as vehicle type.

5.1.11.2 Vehicle characteristics

All econometric methods and AR identified a significant effect of old vehicles (vehicle age ≥ 15) on the most serious crashes. Econometric models provided positive coefficients for both fatal and serious crashes and the results are consistent with AR. The vehicle type involved in the crash with a pedestrian influences the pedestrian outcome. Specifically, a pedestrian struck by a truck has a higher injury risk. The results were highlighted by all methods. A further risk for pedestrian safety was the presence of articulated vehicles, the factor was identified by AR as the strongest two-item rule with fatal crashes. The relation was confirmed by econometric models and RF. By the econometric models, heavy oil vehicles were also identified affecting crash severity with positive coefficients whereas hybrid vehicles exhibited a reduction in crash severity. However, AR also found an association of fatal crashes with vehicles with petrol propulsion. Furthermore, ANN identified vehicle engine capacity affecting pedestrian crash severity.

5.1.11.3 Environmental characteristics

The day of the week, lighting, pavement, and weather at the time of the crash were significant variables. The weekend resulted in a predictor of fatal and serious crashes in both econometric and ML models. In particular, the result of econometric models was confirmed by the association rules. Darkness due to absence of lights or inadequate lighting increase the likelihood of the most severe crashes. Pavement condition affects crash severity particularly when wet or damp. Econometric models and AR found consistent results. Weather conditions were only highlighted by ANN which associates with the weather variable 60% of importance in classification. However, neither the other ML tools nor the econometric models provide this result.

5.1.11.4 Crash characteristics

The number of vehicles involved in the crash played a pivotal role. All econometric models showed an increase in the coefficients for both fatal and serious injuries with multi-vehicle crashes and the relation was also captured by AR (rule 34, L = 5.00). A frontal vehicle impact was identified as critical by AR and the result was confirmed by RF and ANN with the first point of impact having a great influence on the classification process. The number of pedestrians involved affected serious crashes and the association was identified only by AR. The generated two-item rule is the strongest one for serious crashes.

5.1.11.5 Driver characteristics

Gender was among the most important variables exhibited by RF for fatal crashes and the result was consistent with AR and all econometric models which identified males as drivers involved in fatal and serious crashes. Very young drivers (age ≤ 24) were most likely to be involved in the severe crashes. The relation was identified by econometric models (both for fatal and serious crashes) and AR (for serious crashes) and driver age was the most important predictor for ANN. Furthermore, only AR identified aspects related to driver purpose of journey and driver home area. Journey as part of work or commuting to/from work were considered critical both for fatal and serious pedestrian crashes.

5.1.11.6 Pedestrian characteristics

All methods found a correlation between pedestrian age and gender with fatal and serious crashes. Elder pedestrians (at least 65 years old) resulted very exposed to the most serious crashes even though econometric models and AR highlighted pedestrians aged over 75 as the most vulnerable once in a crash. Pedestrian age was also among the strongest predictors in CT, RF, ANN, and SVM variable importance lists with over 50% of the influence on classification. As far as pedestrian gender, only econometric methods and AR found a greater propensity of male pedestrians towards most serious crashes.

5.2 Swedish results

All the explanatory variables reported in the descriptive statistics (in Table 35 and Table 36) were tested for inclusion in the econometric models. The estimation results are reported in Table 80 and Table 81 for the multinomial logit, in Table 84 are provided the results of the mixed multinomial logit, in Table 87 and Table 88 are presented the results related to the ordered logit. No results are provided for the mixed ordered logit as the method does not arrive at convergence. Regarding the machine learning tools, Figure 27 presents the classification tree, Table 95 and Table 96 provide the variable importance for fatal and serious injury classifications in RF, Table 98 and Table 99 present partially the results of AR, Table 103 provides the variable importance for fatal and serious injury classifications in SVM, and Table 107 provides summary results for ANN tool.

Furthermore, the confusion matrix and all the performance metrics evaluated are reported for each method.

5.2.1 Multinomial logit

Statistically significant explanatory variables were 11 and significant indicator variables associated with these categorical variables were 28. 24 significant indicators described the fatal crashes and 22 significant indicators described the serious injury crashes.

Table 80 – Multinomial logit: parameter estimates and goodness of fit measures, Sweden (Part A).

Variable	Fatal				Serious			
	Estimates	OR	Std. err.	P>|z|	Estimates	OR	Std. err.	P>|z|
Intercept	-1.697	0.183	0.206	<0.001	-1.019	0.361	0.152	<0.001
Area (Urban as baseline)								
Rural	0.522	1.685	0.129	<0.001				
Crash Location Detail (Road section as baseline)								
Roundabout	-1.297	0.273	0.317	<0.001				
Interchange	2.971	19.511	0.779	<0.001				
Pedestrian/bicycle path	-0.363	0.696	0.189	0.055	-0.313	0.731	0.106	0.003
Lighting (Daylight as baseline)								
Dawn/dusk	-0.567	0.567	0.204	0.005	-0.651	0.522	0.158	<0.001
Darkness	0.325	1.384	0.085	<0.001	0.181	1.198	0.067	0.007
Speed Limit(30 km/h as baseline)								
40	0.623	1.865	0.145	<0.001	0.544	1.723	0.088	<0.001
50	0.926	2.524	0.128	<0.001	0.191	1.210	0.082	0.020
≥60	2.270	9.679	0.171	<0.001	0.808	2.243	0.135	<0.001
Vehicle Type (Car as baseline)								
Bike	-2.102	0.122	0.326	<0.001	0.437	1.548	0.106	<0.001
PTW	-0.804	0.448	0.296	0.007	0.496	1.642	0.156	0.001
Truck	0.532	1.702	0.104	<0.001	0.299	1.349	0.087	0.001

Table 81 – Multinomial logit: parameter estimates and goodness of fit measures, Sweden (Part B).

Variable	Fatal				Serious			
	Estimates	OR	Std. err.	P>|z|	Estimates	OR	Std. err.	P>|z|
Driver Gender (Male as baseline)								
Female	-0.448	0.639	0.109	<0.001				
Driver Age (25-34 as baseline)								
0-24	-0.328	0.720	0.152	0.031	-0.480	0.706	0.135	0.010
35-44	-0.637	0.529	0.156	<0.001				
45-54	-0.239	0.787	0.144	0.096				
55-64	-1.014	0.363	0.164	<0.001				
65-74	-1.233	0.291	0.192	<0.001	0.227	1.255	0.136	0.095
≥75	-0.538	0.584	0.182	0.003	0.258	1.294	0.146	0.078
Driver Alcohol/Drug use (No as baseline)								
Yes	3.958	52.353	1.072	<0.001	2.710	15.029	1.009	0.007
Pedestrian Gender (Male as baseline)								
Female	-0.562	0.570	0.077	<0.001	-0.098	0.907	0.059	0.098
Pedestrian Age (25-34 as baseline)								
0-14					0.645	1.906	0.123	<0.001
15-24	-0.454	0.635	0.186	0.014				
35-44	0.989	2.689	0.166	<0.001	0.523	1.687	0.129	<0.001
45-54	0.443	1.557	0.174	0.011	0.523	1.687	0.126	<0.001
55-64	1.387	4.003	0.162	<0.001	1.198	3.313	0.122	<0.001
65-74	1.529	4.614	0.167	<0.001	1.447	4.250	0.123	<0.001
≥75	2.560	12.936	0.156	<0.001	1.861	6.430	0.118	<0.001
Vehicle N. Trailers (0 as baseline)								
1	1.981	7.250	0.247	<0.001	2.178	8.829	0.228	<0.001
Log likelihood null model			-10,355.50					
Log likelihood full model			-6,675.74					
R^2McFadden			0.355					

Overall, the MNL model exhibited 26% of correct classification for slight injury crashes, 69% for serious injury, and 64% for fatal crashes with a global accuracy superior to 29%. However, the total correct classification was also evaluated considering the model performances exhibited for the classification of slight injuries, the most frequent class. Thus, even if the global accuracy was very low, serious injury and fatal classification accuracy was more than 60% respectively.

Table 82 – Confusion matrix for the multinomial logit, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	2,324	5,088	1,376
	Serious	20	293	113
	Fatal	6	70	136

Table 83 – Performance of metrics for the multinomial logit model, Sweden.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.323	0.637	0.908
TP_{rate} (Acc+)	0.688	0.642	0.292
Precision	0.054	0.084	0.926
F-measure	0.100	0.148	0.397
G-mean	0.471	0.639	0.501
AUC	0.597	0.845	0.766
Acc	0.292		
Err	0.708		

5.2.2 Mixed multinomial logit

The log-likelihood at zero (-10,355) and at convergence (-5,843) gives a McFadden R^2 of 0.44 which is a very good result. The χ^2 of the LR test is 1,665.32 with 2 degrees of freedom and p-value <0.001, showing that the RPMNL model is superior to the standard MNL model with over 99.9% of confidence.

Two indicator variables showed normally distributed random parameters, with statistically significant standard deviations indicating significant unobserved heterogeneity in the crash data. Considering the normally distributed coefficients, the estimated means and standard deviations of these coefficients provide information on the share of the crashes that places a positive value on the random attribute and the share that places a negative value. The distribution of the coefficient of the speed limit ≤ 30 km/h obtains an estimated mean of -0.475 and estimated standard deviation of 0.653, such that 76.9% of the distribution is below zero and 23.1% above. This results imply that for 76.9% of the observations with speed limits ≤ 30 km/h the probability of the fatal outcome decreased while for 23.1% of the observations the probability of a fatal outcome increased. Similarly, the indicator variable roundabout showed a normal distribution with a mean of -2.474 and a standard deviation of 2.431 for serious injury crashes. This means that for 84.6% of the crashes at roundabout the probability of the severe outcome decreased while for 15.4% of the observations the probability of a severe outcome increased.

Table 84 – Mixed multinomial logit: parameter estimates and goodness of fit measures, Sweden.

Variable	Fatal				Serious			
	Estimate	OR	Std. Err.	P>|z|	Estimate	OR	Std. Err.	P>|z|
Intercept	-1.207	0.299	0.401	0.003	0.015	1.015	0.342	0.966
Area (Urban as baseline)								
Rural	0.632	1.881	0.215	0.003				
Crash Location Detail (Road section as baseline)								
Intersection	-0.505	0.604	0.206	0.014				
Roundabout					**-2.474**	**0.084**	**2.549**	**0.332**
Interchange	1.753	5.772	1.015	0.084				
Pedestrian bicycle path	-0.982	0.375	0.521	0.059				
Lighting (Daylight as baseline)								
Dawn/dusk	0.057	1.059	0.017	0.001	-0.017	0.983	0.526	0.974
Darkness	0.476	1.610	0.169	0.005	-0.268	0.765	0.300	0.371
Speed Limit(40 km/h as baseline)								
≤30	**-0.475**	**0.622**	**0.291**	**0.102**				
50	0.356	1.428	0.147	<0.001	-0.331	0.718	0.333	0.320
≥60	1.751	5.760	0.287	<0.001	0.513	1.670	0.473	0.278
Vehicle Type (Car as baseline)								
Bike	-2.909	0.055	1.360	0.032				
Truck	0.945	2.573	0.178	<0.001				
Driver Gender (Male as baseline)								
Female					-0.099	0.906	0.253	0.695
Driver Age (25-34 as baseline)								
65-74	-0.563	0.569	0.328	0.087				
≥75	-0.599	0.549	0.358	0.094				
Driver Alcohol/Drug (No as baseline)								
Yes	3.281	26.602	1.515	0.030				
Pedestrian Gender (Male as baseline)								
Female	-0.325	0.723	0.155	0.036				
Pedestrian Age (25-34 as baseline)								
0-14	0.510	1.665	0.374	0.173				
15-24	-0.266	0.766	0.374	0.477	-1.036	0.355	0.489	0.034
35-44	0.495	1.640	0.356	0.165				
45-54	0.351	1.420	0.361	0.331				
55-64	1.122	3.071	0.345	0.001	0.605	1.831	0.353	0.087
65-74	1.352	3.865	0.335	<0.001				
≥75	1.998	7.374	0.313	<0.001				
Pedestrian Alcohol/Drug (No as baseline)								
Yes	0.645	1.906	0.391	<0.001	0.964	2.622	0.542	0.076
Standard Deviation of random parameter								
Roundabout					**2.431**	**11.370**	**1.394**	**0.081**
Speed Limit ≤30	**0.653**	**1.921**	**0.383**	**0.088**				
Log likelihood null model			-10,355.50					
Log likelihood full model			-5,843.08					
R^2McFadden			0.436					

The RPMNL model exhibited 62% of correct classification for slight injury crashes, 49% for serious injury, and 93% for fatal crashes with a global accuracy superior to 62%. As for MNL model, the total correct classification was also evaluated considering the model performances exhibited for the classification of slight injuries, the most frequent class. It is noteworthy to observe that the overall accuracy of RPMNL model considerably differs from that of MNL.

Table 85 – Confusion matrix for the mixed multinomial logit, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	5,423	1,157	2,208
	Serious	30	207	189
	Fatal	3	12	197

Table 86 – Performance of metrics for themixed multinomial logit model, Sweden.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.828	0.701	0.818
TP_{rate} (Acc+)	0.486	0.929	0.664
Precision	0.150	0.076	0.383
F-measure	0.230	0.140	0.363
G-mean	0.634	0.807	0.725
AUC	0.494	0.861	0.675
Acc	0.618		
Err	0.382		

5.2.3 Ordered logit

The ordered nature of the response variable crash severity can be explored also in the Swedish database. As for the British estimates, a positive (or negative) parameter implied the likelihood (or unlikelihood) of a severe injury with an increasing value of the explanatory variable and a reduction in the likelihood of a slight injury. Statistically significant explanatory variables were 9 and significant indicator variables associated with these categorical variables were 21 (see Table 87 and Table 88).

Table 87 – Ordered logit: parameter estimates and goodness of fit measures, Sweden (Part A).

Variable	**Estimates**	**OR**	**Std. err.**	**P>|z|**
Intercept	0.445	1.560	0.075	<0.001
Area (Urban as baseline)				
Rural	0.522	1.685	0.129	<0.001
Crash Location Detail (Road section as baseline)				
Interchange	-0.305	0.737	0.054	<0.001
Pedestrian/bicycle path	-0.390	0.677	0.090	<0.001
Lighting (Daylight as baseline)				
Darkness	0.346	1.413	0.050	<0.001
Speed Limit(30 km/h as baseline)				
40	0.310	1.363	0.071	<0.001
50	0.411	1.508	0.066	<0.001
≥60	1.716	5.562	0.094	<0.001
Vehicle Type (Car as baseline)				
Bike	-0.169	0.845	0.091	0.063
Truck	0.845	2.328	0.060	<0.001

Table 88 – Ordered logit: parameter estimates and goodness of fit measures, Sweden (Part B).

Variable	Estimates	OR	Std. err.	P>|z|
Driver Gender (Male as baseline)				
Female	-0.169	0.845	0.058	0.003
Driver Age (25-34 as baseline)				
0-24	-0.417	0.659	0.094	0.031
35-44	-0.588	0.555	0.089	<0.001
45-54	-0.431	0.650	0.085	0.096
55-64	-0.365	0.694	0.088	<0.001
65-74	-0.686	0.504	0.098	<0.001
Pedestrian Gender (Male as baseline)				
Female	-0.379	0.685	0.044	0.044
Pedestrian Age (25-34 as baseline)				
0-14	0.441	1.554	0.099	<0.001
45-54	0.384	1.468	0.099	<0.001
55-64	0.883	2.418	0.094	<0.001
65-74	1.226	3.408	0.094	<0.001
≥75	1.817	6.153	0.089	<0.001
Cut points	-0.406	0.666	0.118	
cut1	1.538	4.655	0.119	
cut2	-0.406	0.666	0.118	
Log likelihood null model		-10,355.50		
Log likelihood full model		-8,391.94		
R^2McFadden		0.190		

Overall, the OL model exhibited 66% of correct classification for slight injury crashes, 59% for serious injury, and 70% for fatal crashes with a global accuracy superior to 65%. In this case study, the OL model exhibited fair global accuracy.

Table 89 – Confusion matrix for the ordinal logit, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	5,349	154	5,349
	Serious	297	5,492	297
	Fatal	164	2,510	164

Table 90 – Performance of metrics for the ordinal logit model, Sweden.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.686	0.833	0.650
TP_{rate} (Acc+)	0.362	0.675	0.599
Precision	0.058	0.114	0.910
F-measure	0.100	0.196	0.706
G-mean	0.498	0.749	0.623
AUC	0.594	0.889	0.749
Acc	0.599		
Err	0.401		

5.2.4 Mixed ordered logit

The results of the mixed ordered logit were not reported in this thesis as the results of the model did not arrive at convergence. The issue is typical with small size samples. However, some previous studied recognized that even in large samples, the non-convergence issue can be a consequence of the frequency distribution of either the dependent or independent variables (Allison, 2008).

5.2.5 Classification tree

The Classification tree obtained for Sweden is reported in Figure 26.

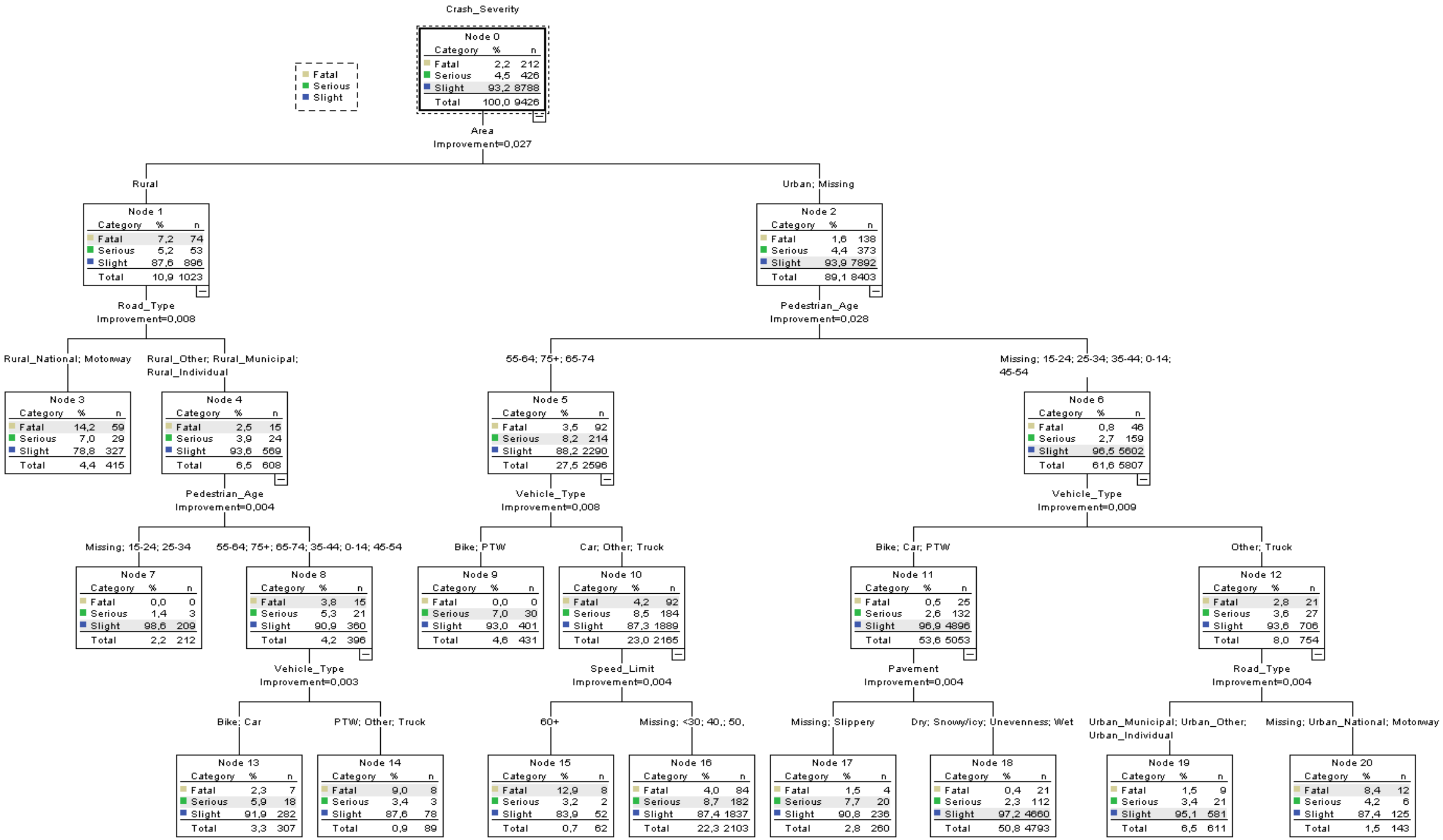


Figure 27 – Classification tree, Sweden.

The tool generated 8 terminal nodes, 3 of which predicted fatal crashes, 2 predicted serious crashes, and 3 predicted slight injury crashes.

The posterior classification ratio (PCR) was assessed for all the nodes (for full details, see APPENDIX 2 ~ SWEDEN, classification tree section) but in the Table 91 PCR values were only reported for the terminal nodes to understand how representative each terminal node is in relation to the predicted class. Node 3 and node 14 exhibited high PCR (6.32 and 4.00 respectively) which implies the robustness of both the terminal nodes for fatal classification.

Table 91 – Terminal nodes and relative Posterior Classification Ratio value, Sweden.

	PCR			
Terminal Nodes	**Fatal**	**Serious**	**Slight**	**Actual Predicted Class**
3	6.32	1.55	0.85	Fatal
7	0.00	0.31	1.06	Slight
13	1.01	1.30	0.99	Serious
14	4.00	0.75	0.94	Fatal
17	0.68	1.70	0.97	Serious
18	0.19	0.52	1.04	Slight
19	0.65	0.76	1.02	Slight
20	3.73	0.93	0.94	Fatal

The analysis of variable importance (Table 92 and Figure 28) identified three variables influencing the classification accuracy of pedestrian crash severity by at least 50%: (1) road type, (2) speed limit, and (3) pedestrian age. Area and vehicle type influenced the classification by more than 40%, followed by pavement, pedestrian gender, number of pedestrians and vehicles involved, lighting, crash location and day of the week.

Table 92 – Independent Variable Importance, Sweden.

Independent Variable	Importance	Normalized Importance
Road Type	0.058	100.0%
Speed Limit	0.053	91.1%
Pedestrian Age	0.034	58.8%
Area	0.027	46.8%
Vehicle Type	0.027	46.0%
Pavement	0.012	20.2%
Pedestrian Gender	0.004	6.2%
N. Pedestrian involved	0.002	3.7%
N. Vehicle involved	0.001	2.1%
Lighting	0.001	1.6%
Crash Location	0.001	1.0%
Day of Week	<0.001	0.6%

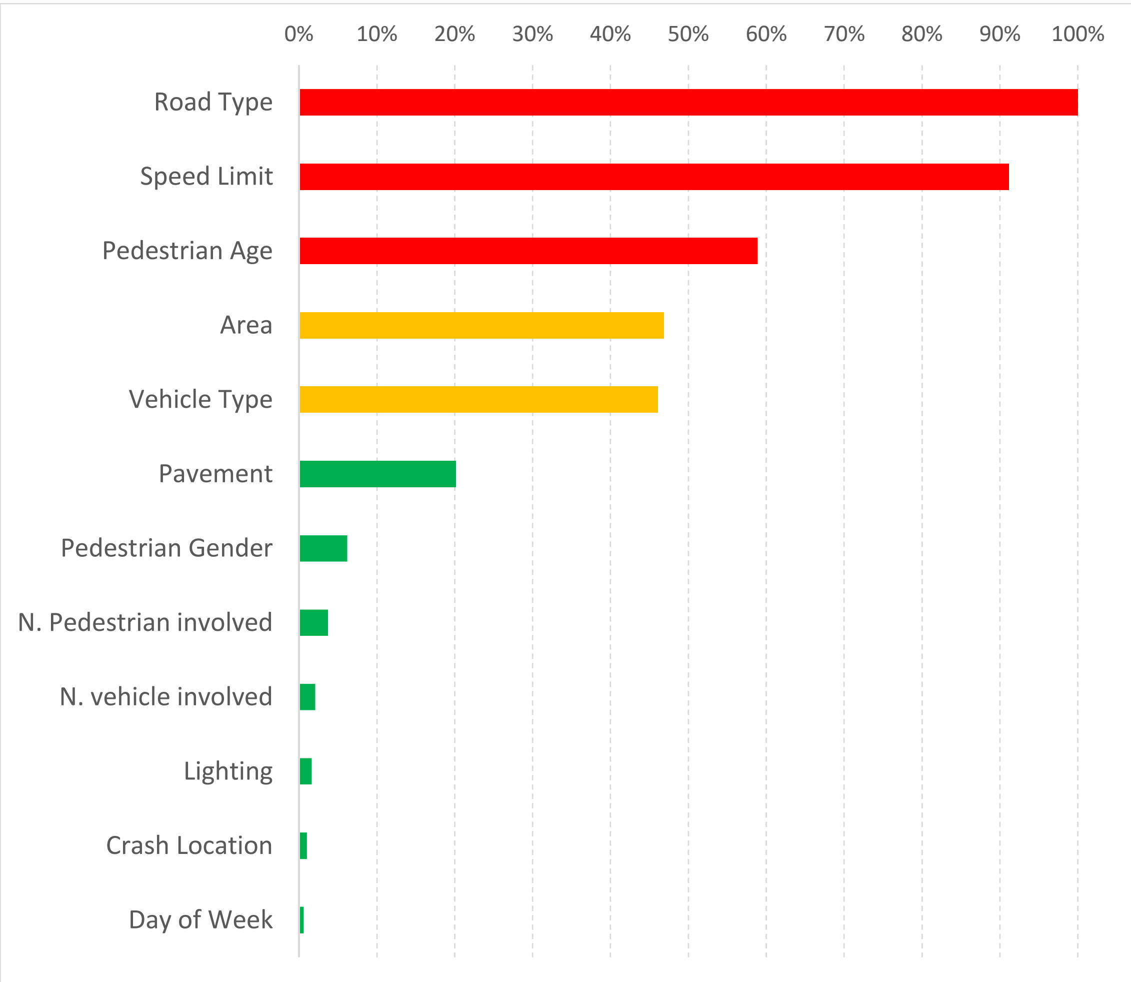


Figure 28 – CT variable normalized importance, Sweden.

Overall, the CT tool exhibited 62% of correct classification for slight injury crashes, 59% for serious injury, and 41% for fatal crashes with a global accuracy superior to 61%.

Table 93 – confusion matrix for the CT, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	5,450	2,756	582
	Serious	136	250	40
	Fatal	30	95	87

Table 94 – Perfromance metrics for the CT, Sweden.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^{-})$	0.660	0.902	0.675
TP_{rate} (Acc^{+})	0.587	0.410	0.614
Precision	0.081	0.123	0.911
F-measure	0.142	0.189	0.716
G-mean	0.622	0.608	0.643
AUC	0.625	0.711	0.808
Acc	0.614		
Err	0.386		

5.2.6 Random forest

Initially, RF was implemented generating 500 trees. However, after setting the optimal number of trees based on the out of bag (OOB) sample error rate, the RF tool was performed with 105 trees. After determining the number of optimal trees, RF was performed again to determine the list of the most important variables associated with pedestrian crash severity in Sweden. As for Great Britain, also for Swedish it was possible to estimate the importance of the predictor variables for fatal and serious crashes distinctly. The importance of each explanatory variable showed driver gender, pedestrian gender, vehicle number of trailers as the variables contributing most to pedestrian crash severity.

All 105 trees were extracted. However, below only the first tree generated by the random forest has been reported. The other trees were reported in the APPENDIX 2 ~ SWEDEN (random forest section).

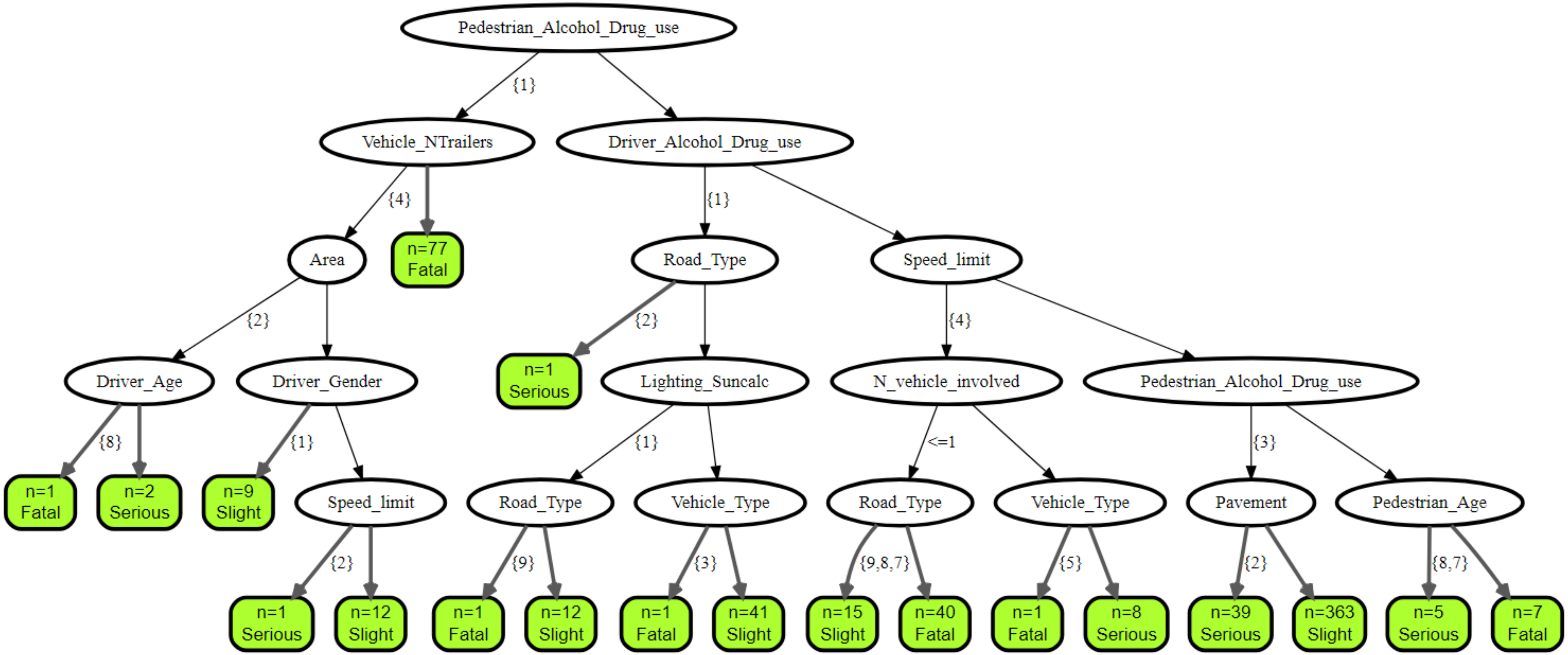


Figure 29 – First tree of the Random Forest, Sweden.

Table 95 – RF Independent Variable Importance for fatal classification, Sweden.

Fatal Independent Variables	Importance	Normalized importance
Driver Gender	0.103	100.0%
Pedestrian Gender	0.095	91.5%
Vehicle N. Trailers	0.086	83.1%
Vehicle Type	0.049	47.2%
Crash Location Detail	0.039	38.2%
Pavement	0.028	27.0%
Pedestrian Age	0.027	25.8%
Pedestrian Alcohol Drug use	0.023	22.2%
Driver Alcohol Drug use	0.018	17.1%
Driver Age	0.015	14.1%
Road Type	0.013	12.4%
Lighting	0.007	7.1%
Speed Limit	0.005	5.3%
N. vehicle involved	0.005	5.0%
Area	0.004	4.0%
Crash Location	0.001	0.9%
N. Pedestrian involved	-0.006	-5.9%
Driver Alcohol Drug use	-0.069	-67.1%

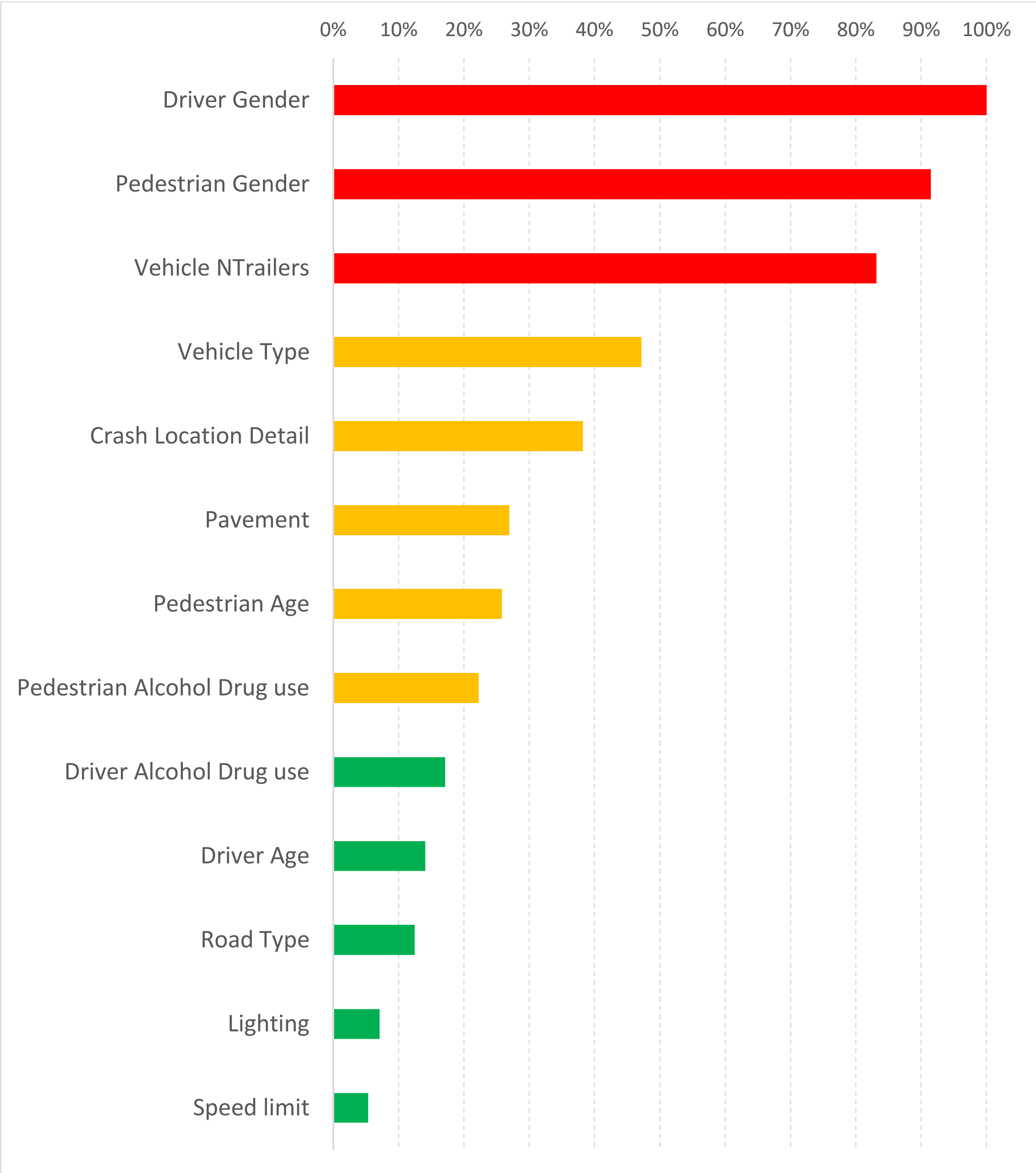


Figure 30 – RF variable normalized importance for fatal classification, Sweden.

Table 96 – RF Independent Variable Importance for serious injury classification, Sweden.

Serious Independent variables	Importance	Normalized importance
Pedestrian Alcohol/Drug use	0.314	100.0%
Pedestrian Age	0.108	34.6%
Driver Alcohol/Drug use	0.103	33.0%
Speed limit	0.052	16.7%
Area	0.039	12.3%
Driver Gender	0.037	11.9%
Lighting	0.026	8.4%
Pedestrian Gender	0.018	5.7%
Crash Location Detail	0.012	3.8%
Driver Age	0.000	-0.1%
Crash Location	-0.003	-0.8%
N. Pedestrian involved	-0.017	-5.3%
Road Type	-0.020	-6.2%
N. vehicle involved	-0.035	-11.1%
Vehicle N. Trailers	-0.045	-14.3%
Vehicle Type	-0.050	-15.9%
Pavement	-0.093	-29.8%

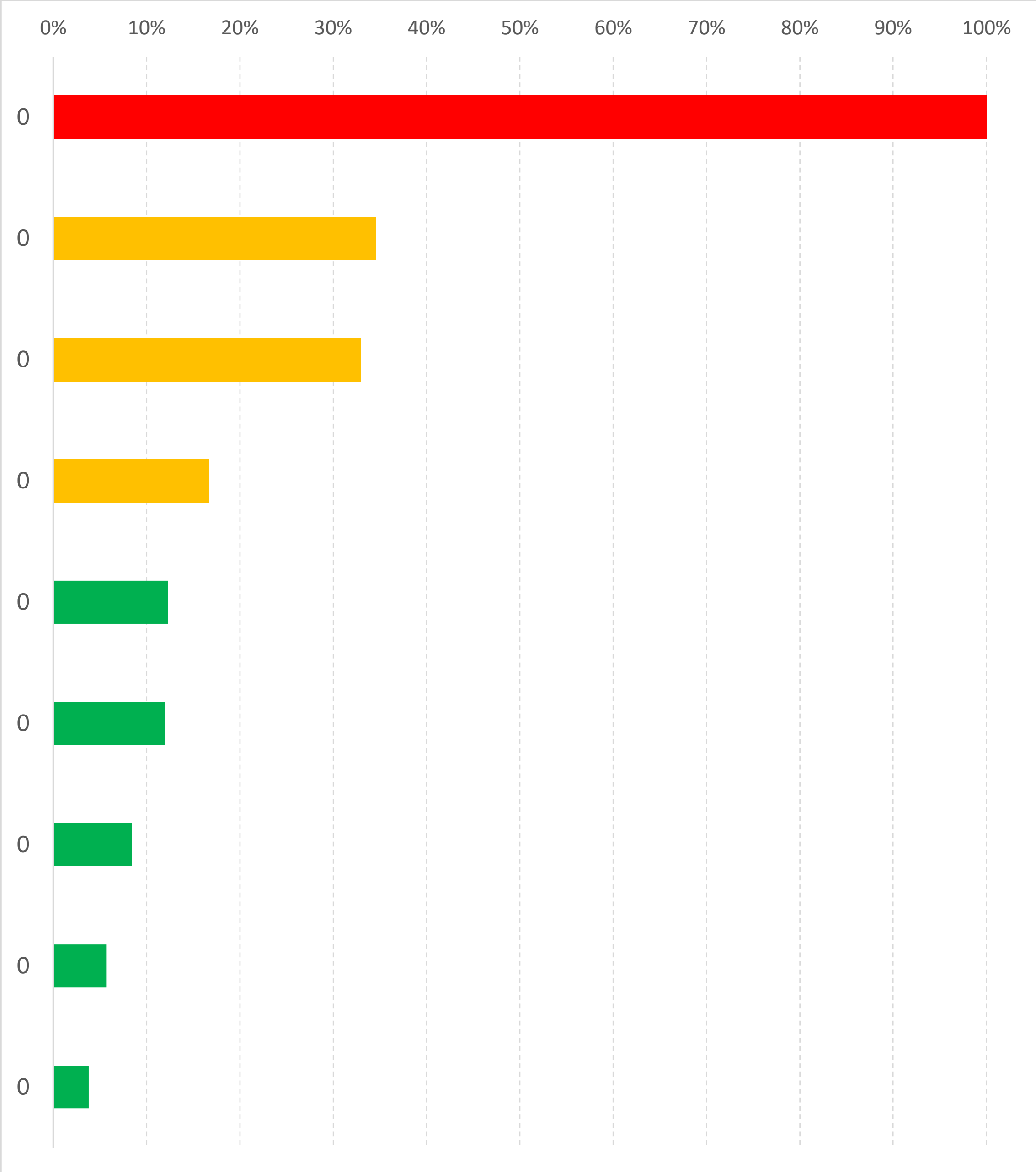


Figure 31 – RF variable normalized importance for serious injury classification, Sweden.

Overall, the RF tool exhibited 83% of correct classification for slight injury crashes, 39% for serious injury, and 59% for fatal crashes with a global accuracy superior to 80%. As expected, also in the Swedish case study, the global accuracy of RF tool was superior to the global accuracy exhibited by CT algorithm.

Table 97 – Confusion matrix for the RF, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	7,267	1,339	182
	Serious	237	168	21
	Fatal	36	51	125

Table 98 – Performance metrics for the RF, Sweden.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^-)$	0.963	0.948	0.640
TP_{rate} (Acc^+)	0.167	0.679	0.874
Precision	0.033	0.239	0.906
F-measure	0.169	0.463	0.881
G-mean	0.576	0.758	0.736
AUC	0.787	0.953	0.091
Acc	0.906		
Err	0.094		

5.2.7 Association rules

The threshold values of support (S), confidence (C), and lift (L) were set as follows: S ≥ 0.2%, C ≥ 3%, L ≥ 1.5, and LIC ≥ 1.05.

The a priori algorithm generated 482 rules with fatal crash as consequent and 105 rules with serious crash as consequent. Furthermore, the extracted rules exhibited up to five items as antecedents. Among the rules with fatal crash as consequent, 381 rules were generated by roadway characteristics. In detail 127 rules included road type (motorway and national roads) as first antecedent, 155 rules included speed limit equal to or greater than 60 km/h, 97 rules included rural area. The remaining rules included pedestrian aged 75 years old and over (27 rules) and truck vehicles (29 rules) as first antecedents. Among the 105 rules with serious injury as consequent, Pedestrian age generated a considerable number of significant rules for serious injury as consequent (76 rules out of 105, over than 70% of the total rules). 19 rules included very old drivers (driver age equal to or greater than 75 years old) followed by 10 rules with urban national roads. As for the rules extracted in the previous section (British database), also in this section, 2-item rules were ordered by the decreasing value of lift, the 3-item rules having the same antecedent of the 2-item rule were ordered again by the decreasing value of the lift, and so on and then the rules were grouped according to the strongest 2-item parent rules. Table 98 and Table 99 contain the strongest rules predicting fatal and serious crashes.

Table 99 – Association rules with fatal as consequent, Sweden.

Antecedents	S %	C %	Lift	LIC
Road Type=Motorway	0.21	24.39	10.84	n.a.
Road Type=Motorway & Driver Alcohol/Drug use=No	0.21	26.32	11.70	1.08
Road Type=Motorway & Driver Alcohol/Drug use=No & N. Pedestrian involved=1	0.21	28.17	12.52	1.07
Road Type=Rural National	0.48	12.82	5.70	n.a.
Road Type=Rural National & Lighting=Darkness	0.31	20.57	9.14	1.60
Road Type=Rural National & Lighting=Darkness & Driver Gender=Male	0.24	28.75	12.78	1.40
Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Day of Week=Weekday	0.20	42.22	18.77	1.47
Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Day of Week=Weekday & Speed Limit≥60	0.20	46.34	20.60	1.64
Road Type=Rural National & Lighting=Darkness & Pavement=Dry	0.30	23.73	10.55	1.15
Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Crash Location Detail=Road section	0.29	25.96	11.54	1.09
Road Type=Rural National & Lighting=Darkness & Speed Limit≥60	0.30	22.05	9.80	1.07
Road Type=Rural National & Lighting=Darkness & Speed Limit≥60 & Crash Location Detail=Road section	0.28	23.64	10.51	1.07
Road Type=Rural National & Pedestrian Gender=Male	0.34	15.46	6.87	1.21
Road Type=Rural National & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.34	17.39	7.73	1.13
Speed Limit≥60	0.82	10.49	4.66	n.a.
Speed Limit≥60 & Vehicle Type=Truck	0.29	22.88	10.17	2.18
Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male	0.28	28.57	12.70	1.25
Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male & Area=Rural	0.20	32.20	14.32	1.13
Speed Limit≥60 & Vehicle Type=Truck & Pedestrian Gender=Male & Pavement=Dry	0.21	26.67	11.86	1.05
Speed Limit≥60 & Driver Gender=Male	0.63	16.08	7.15	1.53
Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness	0.32	21.43	9.53	1.33
Speed Limit≥60 & Lighting=Darkness	0.43	14.80	6.58	1.41
Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pavement=Dry	0.36	23.45	10.43	1.15
Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pavement=Dry & Driver Gender=Male	0.27	30.49	13.56	1.30
Area=Rural	0.79	7.23	3.22	n.a.
Area=Rural & Vehicle Type=Truck	0.23	15.71	6.99	2.17
Pedestrian Age≥75	0.67	5.92	2.63	n.a.
Pedestrian Age≥75 & Driver Gender=Male	0.47	9.32	4.14	1.57
Vehicle Type=Truck	0.67	5.53	2.46	n.a.
Driver Age=25-34	0.45	5.43	2.41	n.a.
Driver Gender=Male	1.62	4.16	1.85	n.a.
Driver Gender=Male & Lighting=Darkness	0.60	5.54	2.46	1.33
Driver Age=45-54	0.34	3.74	1.66	n.a.
Driver Age=45-54 & Crash Location Detail=Road section	0.28	5.15	2.29	1.38
Driver Age=0-24	0.27	3.47	1.54	n.a.

Table 100 – Association rules with serious injury as consequent, Sweden.

Antecedents	S %	C %	Lift	LIC
Pedestrian Alcohol/Drug use=Yes	0.27	25.00	5.53	n.a.
Pedestrian Alcohol/Drug use=Yes & Day of Week=Weekday	0.20	33.93	7.51	1.36
Pedestrian Alcohol/Drug use=Yes & Pedestrian Gender=Male	0.22	29.58	6.54	1.18
Pedestrian Alcohol/Drug use=Yes & Crash Location Detail=Road section	0.20	26.76	5.92	1.07
Pedestrian Alcohol/Drug use=Yes & Area=Urban	0.22	26.58	5.88	1.06
Pedestrian Age≥75	1.12	9.96	2.20	n.a.
Pedestrian Age≥75 & Lighting=Darkness	0.23	13.66	3.02	1.37
Pedestrian Age≥75 & Lighting=Darkness & Day of Week=Weekday	0.20	14.84	3.28	1.09
Pedestrian Age≥75 & Speed Limit=50	0.46	13.27	2.94	1.33
Pedestrian Age≥75 & Speed Limit=50 & Crash Location Detail=Road section	0.29	15.00	3.32	1.13
Pedestrian Age≥75 & Speed Limit=50 & Day of Week=Weekday	0.43	14.96	3.31	1.13
Pedestrian Age≥75 & Crash Location=At intersection	0.32	12.77	2.82	1.28
Pedestrian Age≥75 & Driver Gender=Female	0.25	12.63	2.79	1.27
Pedestrian Age≥75 & Pedestrian Gender=Male	0.43	11.14	2.47	1.12
Pedestrian Age≥75 & Pedestrian Gender=Male & Driver Gender=Male	0.23	12.87	2.85	1.15
Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban	0.37	12.03	2.66	1.08
Pedestrian Age≥75 & Crash Location Detail=Intersection	0.24	10.90	2.41	1.09
Pedestrian Age≥75 & Driver Gender=Male	0.53	10.59	2.34	1.06
Pedestrian Age≥75 & Driver Gender=Male & Speed Limit=50	0.21	12.99	2.87	1.23
N. vehicle involved=2	0.25	9.45	2.09	n.a.
Pedestrian Age=65-74	0.72	8.05	1.78	n.a.
Pedestrian Age=65-74 & Driver Gender=Female	0.21	15.15	3.35	1.88
Pedestrian Age=65-74 & Speed Limit=40	0.22	12.07	2.67	1.50
Pedestrian Age=65-74 & Speed Limit=50	0.25	9.23	2.04	1.15
Pedestrian Age=65-74 & Speed Limit=50 & Day of Week=Weekday	0.23	10.38	2.30	1.12
Pedestrian Age=65-74 & Area=Urban	0.62	8.45	1.87	1.05
Road Type=Urban National	0.33	7.43	1.64	n.a.
Driver Age≥75	0.34	7.13	1.58	n.a.
Driver Age=45-54	0.62	6.78	1.50	n.a.
Driver Age=45-54 & Lighting=Darkness	0.20	7.98	1.77	1.18

Overall, the AR tool exhibited 98% of correct classification for slight injury crashes, 1% for serious injury, and 0% for fatal crashes with a global accuracy superior to 33%. The accuracy reached by AR for the classes of interest was a very low value. This may be the consequence of a too small sample size.

Table 101 – Confusion matrix for the AR, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	2,751	45	6
	Serious	2,718	243	84
	Fatal	3,319	138	122

Table 102 – Performance metrics for the AR, Sweden.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^-)$	0.940	0.971	0.689
TP_{rate} (Acc^+)	0.080	0.034	0.331
Precision	0.570	0.575	0.496
F-measure	0.140	0.064	0.211
G-mean	0.274	0.182	0.228
AUC	0.388	0.804	0.613
Acc	0.331		
Err	0.669		

5.2.8 Support vector machine

SVM model was performed with RBF kernel function. The model returned 7,875 support vectors defining the complex hyperplane. Among the output of the tool, SVM provides the visualization of the most relevant features through non-linear kernels necessary to carry out the classification process. To compare SVM output with the outputs of the other machine learning algorithms implemented in the study, we reported this visualization of the most important predictors exhibited by the tool. SVM identified 5 predictors mostly contributing to the correct classification of pedestrian crash severity: road type, pedestrian age, crash location detail, driver age, and pavement.

Table 103 – SVM Independent Variable Importance, Sweden.

Independent variables	**Importance**	**Normalized Importance**
Road Type	0.111	100.0%
Pedestrian Age	0.111	100.0%
Crash Location Detail	0.099	88.9%
Driver Age	0.086	77.8%
Pavement	0.074	66.7%
Speed limit	0.062	55.6%
Vehicle Type	0.062	55.6%
Lighting	0.049	44.4%
Vehicle N. Trailers	0.037	33.3%
Pedestrian Alcohol Drug use	0.037	33.3%
Driver Alcohol Drug use	0.037	33.3%
Driver Gender	0.037	33.3%
Pedestrian Gender	0.037	33.3%
Crash Location	0.037	33.3%
N. Pedestrian involved	0.037	33.3%
Area	0.025	22.2%
Day of week	0.025	22.2%
N. vehicle involved	<0.001	0.0%

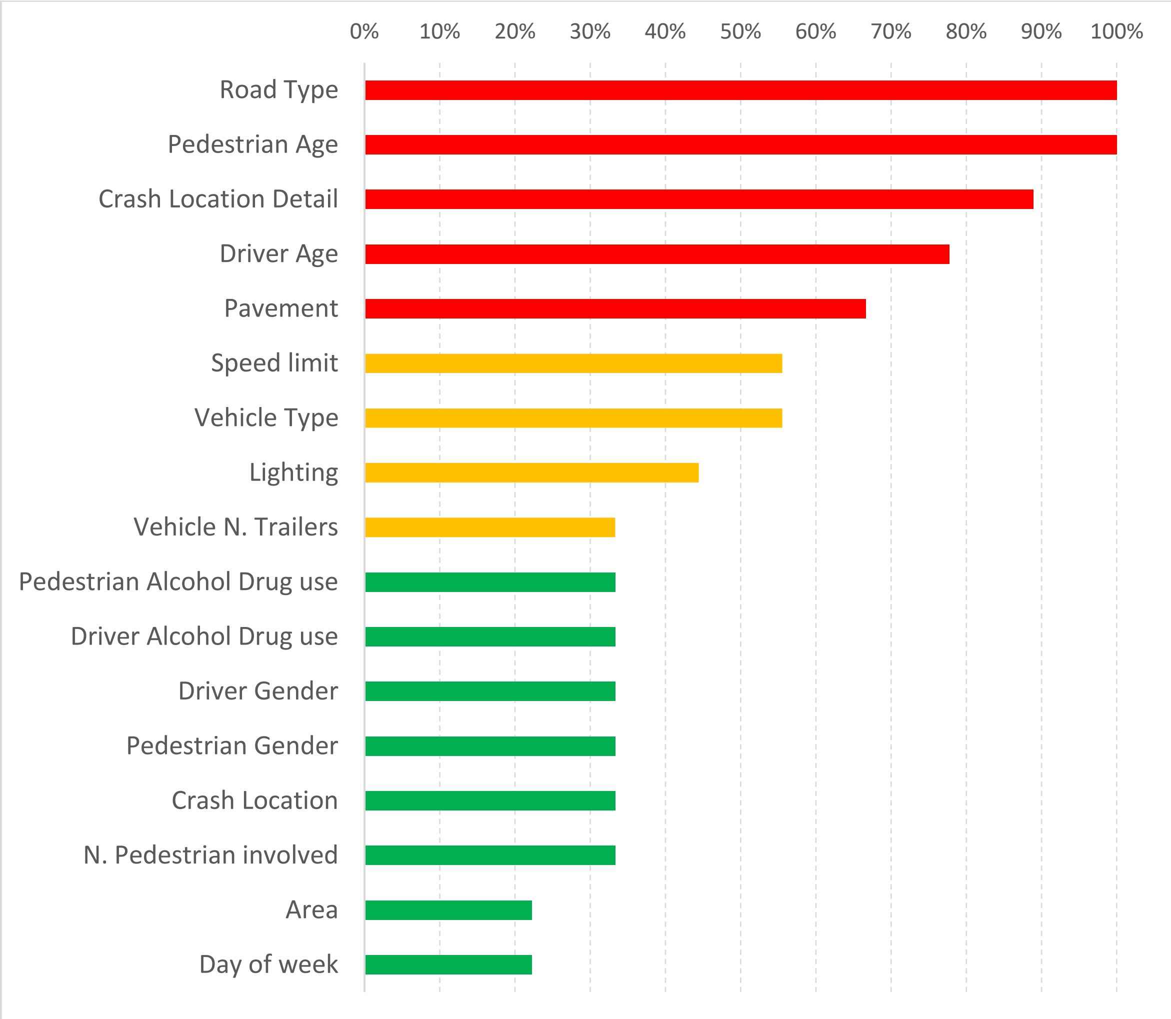


Figure 32 – SVM variable importance, Sweden.

Overall, the SVM tool exhibited 99% of correct classification for slight injury crashes, 73% for serious injury, and 100% for fatal crashes with a global accuracy equal to 74%. SVM reached the highest classification accuracy also in the Swedish case study.

Table 104 – Confusion matrix for SVM, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	8,654	113	21
	Serious	117	309	0
	Fatal	61	0	151

Table 105 – Performance metrics for SVM, Sweden.

	Serious	**Fatal**	**Total**
$TN_{rate}(Acc^{-})$	0.987	0.998	0.739
TP_{rate} (Acc^{+})	0.725	0.712	0.967
Precision	0.732	0.878	0.966
F-measure	0.729	0.786	0.966
G-mean	0.846	0.843	0.843
AUC	0.537	0.857	0.692
Acc	0.967		
Err	0.033		

5.2.9 Artificial neural network

The ANN tool provides a network made up by 18 factors (excluding the bias) namely day of week, area, road type, crash location, crash location detail, pavement, lighting, speed limit, n. vehicle involved, vehicle type, driver gender, driver age, driver alcohol/drug use, vehicle number of trailers, number of pedestrian involved, pedestrian gender, pedestrian age, and pedestrian alcohol/drug use.

Table 106 – Artificial Neural Network general information, Sweden.

Network Information			
Input Layer	Factors	1	Day of Week
		2	Area
		3	Road Type
		4	Crash Location
		5	Crash Location Detail
		6	Pavement
		7	Lighting
		8	Speed Limit
		9	N vehicle involved
		10	Vehicle Type
		11	Driver Gender
		12	Driver Age
		13	Driver Alcohol/Drug use
		14	Vehicle N Trailers
		15	N Pedestrian involved
		16	Pedestrian Gender
		17	Pedestrian Age
		18	Pedestrian Alcohol/Drug use
	Number of Units[a]		81
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		5
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Crash Severity
	Number of Units		3
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

Overall, 81 predictor indicators were connected with 1 unit belonging to the hidden layer through the hyperbolic tangent activation function. Then, through the softmax function, the input layer predictor was finally linked with the output layer represented by the crash severity variable which, based on Swedish data, in organized on three levels of severity. The parameter estimates provided by the tool were reported in the appendix (the reader refers to APPENDIX 2 ~ SWEDEN, Artificial Neural Network section).

Table 107 – ANN Independent Variable Importance, Sweden.

Independent variables	**Importance**	**Normalized Importance**
Pedestrian Alcohol/Drug use	0.147	100.0%
Vehicle Type	0.098	66.9%
Vehicle N. Trailers	0.079	54.1%
Pedestrian Age	0.079	53.6%
Pedestrian Gender	0.076	51.7%
Driver Age	0.073	49.9%
Speed Limit	0.072	49.0%
Road Type	0.066	45.1%
Crash Location Detail	0.065	44.0%
Pavement	0.054	36.8%
Lighting	0.048	32.9%
Driver Alcohol/Drug use	0.030	20.3%
N. vehicle involved	0.028	19.2%
N. Pedestrian involved	0.023	15.9%
Driver Gender	0.021	14.0%
Area	0.020	13.3%
Day of Week	0.015	10.2%
Crash Location	0.007	4.8%

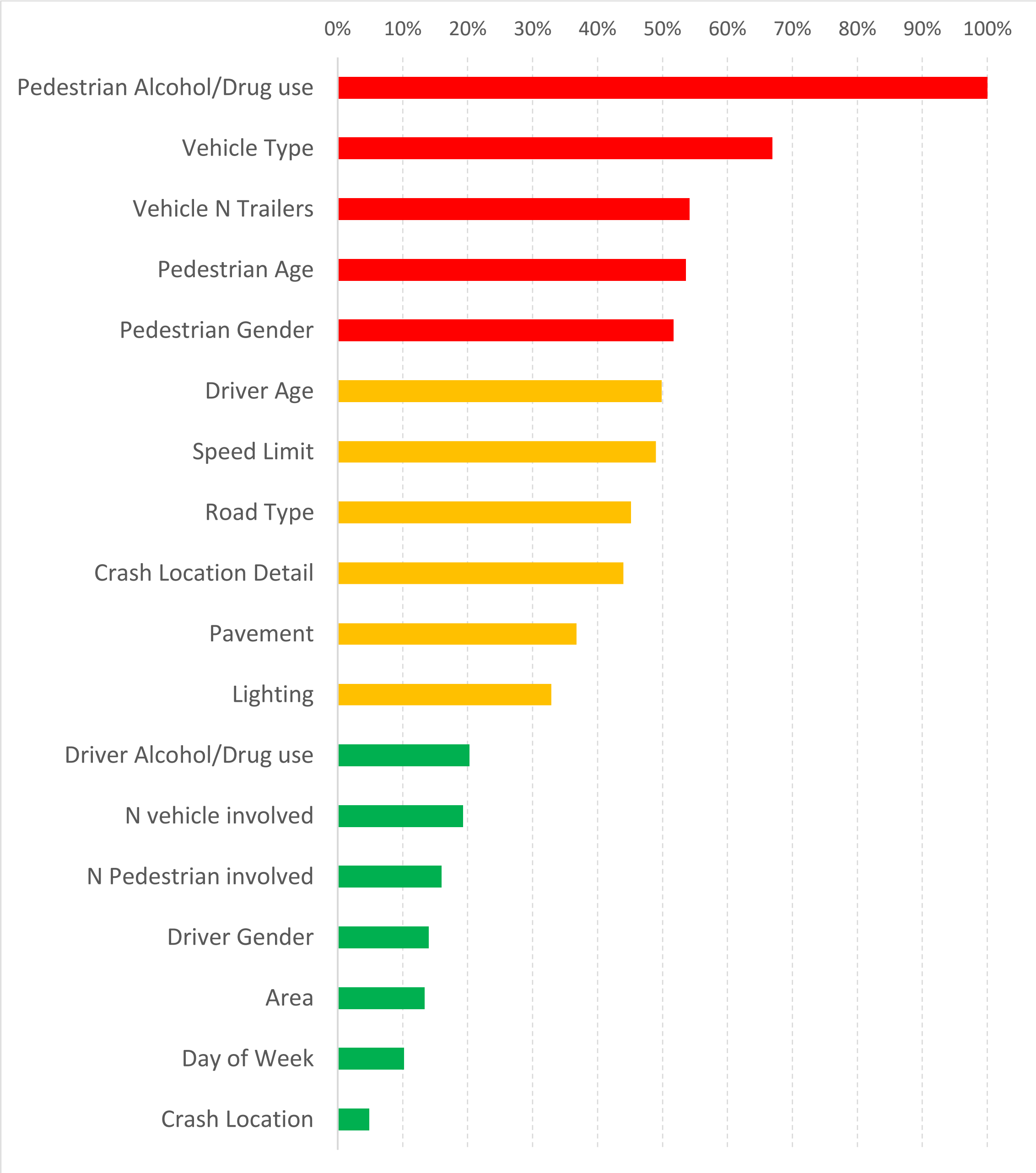


Figure 33 – ANN variable importance, Sweden.

Overall, the ANN tool exhibited 57% of correct classification for slight injury crashes, 48% for serious injury, and 71% for fatal crashes with a global accuracy equal to 61%. The global accuracy was equal to the accuracy reached in British case study.

Table 108 – Confusion matrix for the ANN, Sweden.

		Predicted		
		Slight	Serious	Fatal
Observed	Slight	4,974	3,606	207
	Serious	161	205	58
	Fatal	11	50	150

Table 109 – Performance metrics for the ANN, Sweden.

	Serious	**Fatal**	**Total**
TN_{rate}(Acc-)	0.584	0.951	0.788
TP_{rate} (Acc+)	0.483	0.711	0.611
Precision	0.053	0.361	0.262
F-measure	0.096	0.479	0.332
G-mean	0.531	0.822	0.694
AUC	0.671	0.913	0.808
Acc	0.566		
Err	0.434		

5.1.10 Synthesis of the results

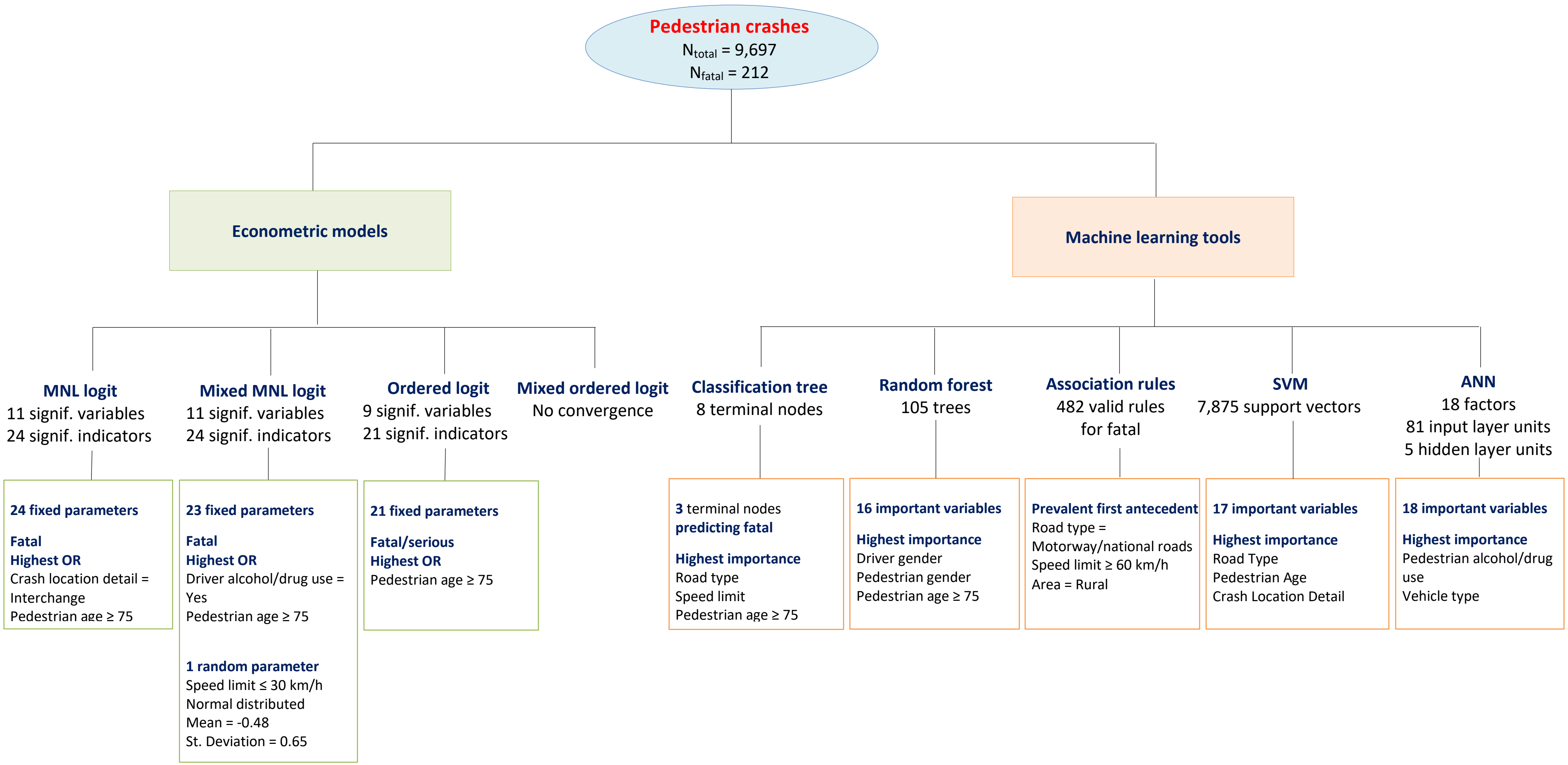


Figure 34 – Main results of the econometric models and machine learning tools for fatal crashes, Sweden.

5.2.11 Measures of performance

As for the British case study, models' performances were evaluated by the F-measure, G-mean, and AUC. Given that the implementation of the weighted approach improves all the methods classification performances (as demonstrated for the British dataset), the measures of performance exhibited by the models were provided considering the weighted formulation. Results of the methods are shown in Table 110.

Table 110 – Measures of performance of weighted econometric models and machine learning algorithms, Sweden.

	Econometric models					Machine learning				
	MNL	RPMNL	OL	RPOL		AR	CT	RF	ANN	SVM
Fatal										
F-measure	0.15	0.14	0.20			0.06	0.19	0.46	0.48	0.79
G-mean	0.64	0.81	0.75	*		0.18	0.61	0.76	0.82	0.86
AUC	0.85	0.86	0.89			0.80	0.71	0.95	0.91	0.86
Serious										
F-measure	0.10	0.23	0.10			0.14	0.14	0.17	0.10	0.73
G-mean	0.47	0.63	0.50	*		0.27	0.62	0.58	0.53	0.85
AUC	0.60	0.49	0.59			0.39	0.63	0.79	0.67	0.54
Averaged performances										
F-measure	0.12	0.20	0.11			0.11	0.16	0.27	0.22	0.75
G-mean	0.53	0.69	0.62	*		0.24	0.62	0.64	0.63	0.85
AUC	0.68	0.62	0.69			0.53	0.65	0.84	0.75	0.64

**The RPOL did not arrive at convergence.*

The performances of the methods applied to the Swedish database reveal that SVM confirmed the higher performances in classifying both fatal and serious injury crashes.

As far as the econometric models are concerned, differently from the British case study, MNL (fixed parameters) and RPMNL (mixed parameters) models did not exhibit better classification performances compared with the ordered version (OL). The results of the methods applied on the Swedish database find out that, among all the econometric models implemented, OL has the best predictive performances. However, the model provides fewer significant variables and indicators than the other econometric models.

As far as the machine learning tools are concerned, SVM outperformed the other algorithms reaching accuracy in both correct positive and negative case classification equal to 86% and the model's good overall performance assessed considering also the minority class is equal to 79%. SVM was followed by ANN and RF. AR and CT exhibited similar performances. The performances of the models got worst from fatal crashes to severe injury made an exception for SVM.

Overall, machine learning algorithms outperformed econometric models and the best performances were reached by SVM and RF.

5.2.12 Significant explanatory variables and effects on crash severity

In Table 111 and Table 112 the significant explanatory variables associated with an increase in crash severity were summarized. Table 111 contains variables associated with an increase in fatal crash probability while Table 112 contains variables associated with an increase in serious crash probability. As far as fatal crashes are concerned, 13 variables are significant both in the econometric models as well as in the machine learning algorithms and 5 variables are significant only in the machine learning algorithms. No further variables are identified by the econometric models only. The result provides confirmation of what has already been found with the British database on data-driven method's ability to uncover more hidden correlations among data than econometric models. The same variables identified by both the groups of methods implemented for fatal crashes are significant also for the serious injury classification, made an exception for the variable area which has been identified by machine learning algorithms only.

Table 111 – Variables associated with an increase in fatal crash probability, Sweden.

Both econometric/ML models	**Econometric models**	**ML algorithms**
Area	-	Crash Location
Crash Location Detail		Day of Week
Driver Age		N. of pedestrian involved
Driver Alcohol Drug use		N. of vehicle involved
Driver Gender		Pavement
Lighting		
Pedestrian Age		
Pedestrian Alcohol/Drug		
Pedestrian Gender		
Road Type		
Speed Limit		
Vehicle N. Trailers		
Vehicle Type		

Table 112 – Variables associated with an increase in serious injury crash probability, Sweden.

Both econometric/ML models	Econometric models	ML algorithms
Crash Location Detail	-	Area
Driver Age		Crash Location
Driver Alcohol/Drug use		Day of week
Driver Gender		N. of pedestrian involved
Lighting		N. of vehicle involved
Pedestrian Age		Pavement
Pedestrian Alcohol Drug use		
Pedestrian Gender		
Road Type		
Speed Limit		
Vehicle N. Trailers		
Vehicle Type		

5.2.11.1 Roadway characteristics

The area variable was the first split in CT growth process involving also higher speed limits associated with fatal crashes, especially in urban area. On the other hand, the unordered econometric models identified rural area over the increase of the speed limit as contributory factors to the increase in the fatal severity of pedestrian crashes. The OL identified, instead, motorways and different rural roads, over speed limits, contributing to the most serious crashes. AR identified high-lift rules with fatal severity as consequent and speed limit ≥ 60 km/h as antecedent (the two-item rule is the rule 128 with lift=4.66) and the association with rural area increases the probability of fatal crashes by 33% whereas urban area was associated with serious injuries. Lower speed limits (40-50 km/h) were associated with serious pedestrian crashes. However, these values of speed limits did not generate two-item rules but increased the probability of serious crashes when associated with the involvement of elder pedestrians (aged over 75). The speed limit was also identified as one of the most important predictors by SVM and ANN with 55.6% and 49.0% importance respectively. Noteworthy, the speed limit ≤30 km/h was random in the RPMNL model with a slightly more 20% of crashes increasing the fatal severity likelihood.

As far crash location detail variable is concerned and considering road section as a baseline, MNL and RPMNL models estimated negative coefficients for roundabout (for fatal crashes) and pedestrian/bicycle path (both for serious and fatal crashes) meaning that crashes involving pedestrians occurred on paths for pedestrians can reduce the severity. Moreover, roundabout resulted in a random parameter for serious injury prediction assuming at almost 85% of crashes a negative coefficient meaning that crashes at the roundabouts are more likely to result in slight injury. AR further strengthened the results by finding rules associated with the most severe crashes on the road section, not at intersections.

5.2.11.3 Vehicle characteristics

Consistently with GB results, trucks increase the likelihood of fatal and serious crashes whereas pedestrian crashes involving PTWs or bikes are more likely to be slight. The effects of these factors on crash severity were consistent among all methods made an exception for OL. Vehicle type was further the second most important variable in ANN. RF and ANN also identified the number of trailers as the third most important predictor with importance superior to 80% and 50% respectively. According to SVM result, the variable has importance superior to 30%. The result was also found by all econometric models with an increase in fatal and severe crashes in presence of vehicles with at least one trailer.

5.2.11.2 Environmental characteristics

Considering daylight as a baseline, darkness was associated with an increase in crash severity (both fatal and serious injury). The relationship was found by all econometric models and by AR. Nevertheless, AR found an increase in pedestrian crash severity when darkness was associated as antecedent to other two-item rules (i.e., rule 8 for fatal crashes, the association of darkness with rural national roads increases the likelihood of observing fatal crashes by 60%. SVM also found lighting an important predictor in the classification process with 44.4% of normalized importance. As far as dawn/dusk lighting is concerned, results were in contrast. MNL found both for fatal and serious crashes a decrease in the most severe crash probability (fatal and serious) compared with daytime. RPMNL found an increase in fatal crash likelihood whereas a decrease in severe injury was observed. OL estimated a negative coefficient which is unique for fatal and severe crashes.

Day of the week and pavement were identified only by ML tools. In particular, AR found weekdays (Monday-Friday) influencing both fatal and serious injury crashes. The variable exhibited 22.2% importance in classification in SVM and 10.2% in ANN.

Slippery pavement was identified by CT contributing to serious pedestrian crashes (node 17). For serious crashes, the result was not confirmed by RF whereas the variable showed 27.0% importance in fatal classification process.

5.2.11.4 Crash characteristics

Only ML algorithms highlighted the dependence of the number of pedestrians and vehicles involved in the crash with the crash severity itself. AR and SVM found the association of the number of pedestrians involved with fatal and severe crashes whereas the association of the number of the vehicles involved with serious injury was discovered by AR with the two-item rule 77 (lift=2.09) according to which the involvement of 2 vehicles in a crash has a significant impact of serious injuries.

5.2.11.5 Driver characteristics

According to the MNL and OL, male drivers aged between 25-34 years old increase the likelihood of fatal crashes. The result regarding the driver gender was not confirmed by the RPMNL but it was, however, consistent with AR which revealed the association of male drivers and fatal crashes with a two-item rule with a lift equal to 1.85. Driver gender was also the strongest predictor in RF fatal classification. When it comes to serious injury crashes, MNL revealed drivers over 65 years old are more prone to this crash severity level whereas AR found the strongest two-item rule with driver age when the driver is older than 75. The rule 78 exhibited a lift equal to 1.58.

The driver's alcohol/drug use influenced both fatal and serious injury crashes. The result was provided by MNL and OL for both levels of severity whereas for RPMNL it was significant for fatal crashes only and for RF the variable had a great impact on serious injury classification.

5.2.11.6 Pedestrian characteristics

Male pedestrians increased the probability of being involved in fatal and serious crashes. The results were consistent among all methods.

Pedestrian age resulted in the first predictor in SVM, the second predictor in RF serious injury classification process (35% of normalized importance) and the third predictor in CT growth process with an influence of almost 60% on classification. AR revealed with rule 9 (serious injury as consequent, the high association between elder pedestrians (over 75) and serious injury crashes (lift = 2.20) whereas the association with elder pedestrians and fatal crashes was even stronger (rule 380, fatal as consequent, lift = 2.63).

Pedestrian alcohol/drug use was the most important variable in serious injury classification for RF. The variable further generated the strongest rule having serious injury as consequent: rule 1 (serious injury section) exhibited a very high lift equal to 5.53.

5.3 Italian results

All the explanatory variables reported in the descriptive statistics (from Table 37 to Table 39) were tested for inclusion in the econometric models. The estimation results are reported in Table 113 and Table 114 for the binary logit, in Table 117 and Table 118 are provided the results of the mixed binary logit. No results are provided for the ordered logit and mixed ordered logit as the crash severity variable does not allow to consider an order among severity levels. Regarding the machine learning tools, Figure 35 presents the classification tree, Table 125 provides the variable importance for fatal classification in RF, Table 128 presents partially the results of AR, Table 131 provides the variable importance for fatal classification in SVM, and Table 134 provides summary results for ANN tool.

Furthermore, the confusion matrix and all the performance metrics evaluated are reported for each method.

5.3.1 Logit

After excluding statistically not significant variables, 17 variables were retained in the final model (Table 113 and Table 114): day of week, season, road type, lighting, alignment, pavement, weather, vehicle type, driver behaviour, driver psychophysical state, driver age, driver gender, pedestrian behaviour, pedestrian psychophysical state, pedestrian age, and pedestrian gender. Furthermore, significant indicator variables associated with these categorical variables were 51.

Table 113 – Logit: parameter estimates and goodness of fit measures, Italy (Part A).

Variable	Estimate	OR	Std. Error	P>|z|
(Intercept)	-2.175	0.114	0.052	<0.001
Road type (Urban Municipal as baseline)				
Motorway	2.972	19.531	0.099	<0.001
Rural Municipal	1.449	4.259	0.050	<0.001
Rural National	2.131	8.423	0.055	<0.001
Rural Provincial	2.063	7.870	0.044	<0.001
Urban National	0.845	2.328	0.038	<0.001
Urban Provincial	1.019	2.770	0.030	<0.001
Lighting (Day as baseline)				
Night	1.044	2.841	0.019	<0.001
Weather (Clear as baseline)				
Fog	-0.239	0.787	0.093	0.010
High winds	-1.075	0.341	0.300	<0.001
Raining	-0.386	0.680	0.042	<0.001
Pavement (Dry as baseline)				
Snowy/Frozen	-0.659	0.517	0.191	<0.001
Slippery	0.355	1.426	0.151	0.019
Day of Week (Weekday as baseline)				
Weekend	0.306	1.358	0.018	<0.001
Season (Summer as baseline)				
Autumn	-0.297	0.743	0.024	<0.001
Spring	-0.241	0.786	0.026	<0.001
Winter	-0.346	0.708	0.026	<0.001
Alignment (Tangent as baseline)				
No Signalized Intersection	-0.502	0.605	0.020	<0.001
Roundabout	-0.896	0.408	0.065	<0.001
Signalized Intersection	-0.483	0.617	0.038	<0.001
Vehicle Type (Car as baseline)				
PTW	-0.322	0.725	0.027	<0.001
Truck	0.939	2.557	0.028	<0.001

Table 114 – Logit: parameter estimates and goodness of fit measures, Italy (Part B).

Variable	Estimate	OR	Std. Error	P>|z|
Vehicle Defect (No as baseline)				
Yes	0.536	1.709	0.122	<0.001
Driver Age (25-44 as baseline)				
0-17	-0.215	0.807	0.081	0.008
18-24	0.324	1.383	0.028	<0.001
45-54	-0.209	0.811	0.022	<0.001
55-64	-0.106	0.899	0.025	<0.001
65-74	-0.322	0.725	0.029	<0.001
≥75	-0.198	0.820	0.029	<0.001
Driver Gender (Female as baseline)				
Male	0.389	1.476	0.020	<0.001
Driver Behaviour (Normal as baseline)				
Disobeying stop sign	2.565	13.001	0.172	<0.001
Distraction	0.677	1.968	0.082	<0.001
Manoeuvring	-0.641	0.527	0.034	<0.001
Speeding	0.701	2.016	0.028	<0.001
Tailgating	0.343	1.409	0.089	<0.001
Driver Psychophysical State (Normal as baseline)				
Alcohol	1.114	3.047	0.069	<0.001
Dazzled	0.310	1.363	0.080	<0.001
Drug	1.979	7.236	0.121	<0.001
Exceeding the prescribed driving period	2.262	9.602	0.625	<0.001
Illness	2.210	9.116	0.176	<0.001
Sleeping	1.969	7.164	0.273	<0.001
Pedestrian Age (25-44 as baseline)				
0-14	-0.081	0.922	0.041	0.048
15-24	-0.279	0.757	0.037	<0.001
45-54	0.372	1.451	0.031	<0.001
55-64	1.006	2.735	0.030	<0.001
65-74	1.586	4.884	0.030	<0.001
≥75	2.406	11.090	0.027	<0.001
Pedestrian Gender (Female as baseline)				
Male	0.479	1.614	0.016	<0.001
Pedestrian Behaviour (Walking regularly as baseline)				
Walking back to the traffic	0.121	1.129	0.045	0.007
Crossing on pedestrian crossing facility	-0.092	0.912	0.038	0.016
No Crossing on pedestrian crossing facility	0.157	1.170	0.035	<0.001
Pedestrian Psychophysical State (Normal as baseline)				
Alcohol	0.545	1.725	0.118	<0.001
Log likelihood null model		-102,957.30		
Log likelihood full model		-74,509.68		
R^2McFadden		0.276		

Overall, the binary logit model exhibited 26% of correct classification for 76% for injury crashes and 76% for fatal crashes with a global accuracy superior to 76%. The binary logit exhibited very good classification accuracy in correctly classifying fatal crashes.

Table 115 – Confusion matrix for the logit model, Italy.

		Predicted	
		Injury	Fatal
Observed	Injury	74,141	23,922
	Fatal	718	2,251

Table 116 – Performance metrics for the logit model, Italy.

Performance metrics	
TN_{rate}(Acc-)	0.756
TP_{rate} (Acc+)	0.758
Precision	0.086
F-Measure	0.154
G-Mean	0.757
AUC	0.837
Acc	0.756
Err	0.244

5.3.2 Mixed logit

After excluding statistically not significant variables, 17 variables were retained in the final model (Table 117 and Table 118): day of week, season, road type, lighting, alignment, pavement, weather, vehicle type, driver behaviour, driver psychophysical state, driver age, driver gender, pedestrian behaviour, pedestrian psychophysical state, pedestrian age, and pedestrian gender. Associated with these categorical variables, 51 indicator variables were found to be significant.

The log-likelihood at zero (-102,957) and at convergence (-50,404) gives a McFadden R^2 of 0.51 which is actually an excellent result. The χ^2 of the LR test is 48,209.70 with 1 degree of freedom and p-value <0.001, showing that the RPMNL model is superior to the standard MNL model with over 99.9% of confidence.

Night-time was found to vary across observations according to a normal distribution with a mean of 1.07 and deviation of 0.45 (Table 117). This means that for 99.06% of the observations the night-time condition increases the probability of a fatal crash while for 0.94% of the observations it leads to a decrease in that probability.

Table 117 – Mixed logit: parameter estimates and goodness of fit measures, Italy (Part A).

Variable	**Estimate**	**OR**	**Std. Err.**	**P>|z|**
Intercept	-2.208	0.110	0.054	<0.001
Day of Week (Weekday as baseline)				
Weekend	0.314	1.369	0.019	<0.001
Season (Summer as baseline)				
Autumn				<0.001
Spring	-0.298	0.742	0.024	<0.001
Winter	-0.241	0.786	0.026	<0.001
Road type (Urban Municipal as baseline)				
Motorway	3.078	21.715	0.104	<0.001
Rural municipal	1.475	4.371	0.051	<0.001
Rural national	2.164	8.706	0.058	<0.001
Rural provincial	2.094	8.117	0.046	<0.001
Urban national	0.851	2.342	0.038	<0.001
Urban provincial	1.031	2.804	0.031	<0.001
Alignment (Tangent as baseline)				
Unsignalized Intersection	-0.510	0.600	0.021	<0.001
Roundabout	-0.898	0.407	0.066	<0.001
Signalized Intersection	-0.479	0.619	0.038	<0.001
Pavement (Dry as baseline)				
Snowy/Frozen	-0.686	0.504	0.195	<0.001
Slippery	0.347	1.415	0.154	0.02
Weather (Clear as baseline)				
Fog	-0.247	0.781	0.096	0.01
High winds	-1.083	0.339	0.307	<0.001
Raining	-0.403	0.668	0.043	<0.001
Lighting (Day as baseline)				
Mean Night	**1.067**	**2.907**	**0.020**	**<0.001**
SD Night	**0.454**	**1.575**	**0.077**	**<0.001**
Vehicle Type (Car as baseline)				
PTW	-0.333	0.717	0.027	<0.001
Truck	0.952	2.591	0.028	<0.001

Table 118 – Mixed logit: parameter estimates and goodness of fit measures, Italy (Part B).

Variable	Estimate	OR	Std. Err.	P>|z|
Vehicle Defect (No as baseline)				
Yes	0.527	1.694	0.124	<0.001
Driver Behaviour (Normal as baseline)				
Disobeying stop sign	2.661	14.311	0.175	<0.001
Distraction	0.685	1.984	0.083	<0.001
Manoeuvring	-0.654	0.520	0.034	<0.001
Speeding	0.715	2.044	0.029	<0.001
Tailgating	0.340	1.405	0.090	<0.001
Driver Psychophysical State (Normal as baseline)				
Alcohol	1.121	3.068	0.071	<0.001
Dazzled	0.315	1.370	0.081	<0.001
Drug	2.082	8.020	0.126	<0.001
Exceeding the prescribed driving period	2.262	9.602	0.623	<0.001
Illness	2.265	9.631	0.178	<0.001
Sleeping	2.031	7.622	0.275	<0.001
Driver Age (25-44 as baseline)				
0-17	-0.219	0.803	0.082	0.01
18-24	0.327	1.387	0.029	<0.001
45-54	-0.216	0.806	0.023	<0.001
55-64	-0.104	0.901	0.025	<0.001
65-74	-0.330	0.719	0.029	<0.001
≥75	-0.205	0.815	0.030	<0.001
Driver Gender (Female as baseline)				
Male	0.390	1.477	0.020	<0.001
Pedestrian Psychophysical State (Normal as baseline)				
Alcohol	0.586	1.797	0.121	<0.001
Pedestrian Behaviour (Walking regularly as baseline)				
Walking back to the traffic	0.120	1.127	0.045	0.01
Crossing on pedestrian crossing facility	-0.101	0.904	0.039	0.01
Crossing outside pedestrian crossing facility	0.168	1.183	0.036	<0.001
Pedestrian Age (25-44 as baseline)				
0-14	-0.082	0.921	0.041	0.05
15-24	-0.286	0.751	0.038	<0.001
45-54	0.365	1.441	0.032	<0.001
55-64	1.014	2.757	0.032	<0.001
65-74	1.612	5.013	0.032	<0.001
≥75	2.432	11.382	0.030	<0.001
Pedestrian Gender (Female as baseline)				
Male	0.492	1.636	0.016	<0.001
Log likelihood null model		-102,957.30		
Log likelihood full model		-50,404.83		
R^2McFadden		0.510		

The random parameter binary logit model exhibited 76% of correct classification for injury crashes and 77% for fatal crashes with a global accuracy superior to 76%. As for MNL model, the total correct classification was also evaluated considering the model performances exhibited for the classification of injury crashes, the most frequent class. It is noteworthy to observe that the overall accuracy of RPLogit model exhibited the same global accuracy of the binary logit. However, the random parameter model also found randomness among data which cannot be captured by the standard formulation of the logit model.

Table 119 – Confusion matrix for the mixed logit, Italy.

		Predicted	
		Injury	Fatal
Osserved	Injury	74,521	23,542
	Fatal	695	2,274

Table 120 – Performance metrics for the mixed logit, Italy.

Performance metrics	
TN_{rate}(Acc-)	0.760
TP_{rate} (Acc+)	0.766
Precision	0.088
F-Measure	0.158
G-Mean	0.763
AUC	0.845
Acc	0.760
Err	0.240

5.3.3 Classification tree

The Classification tree obtained for Great Britain is reported in Figure 35.

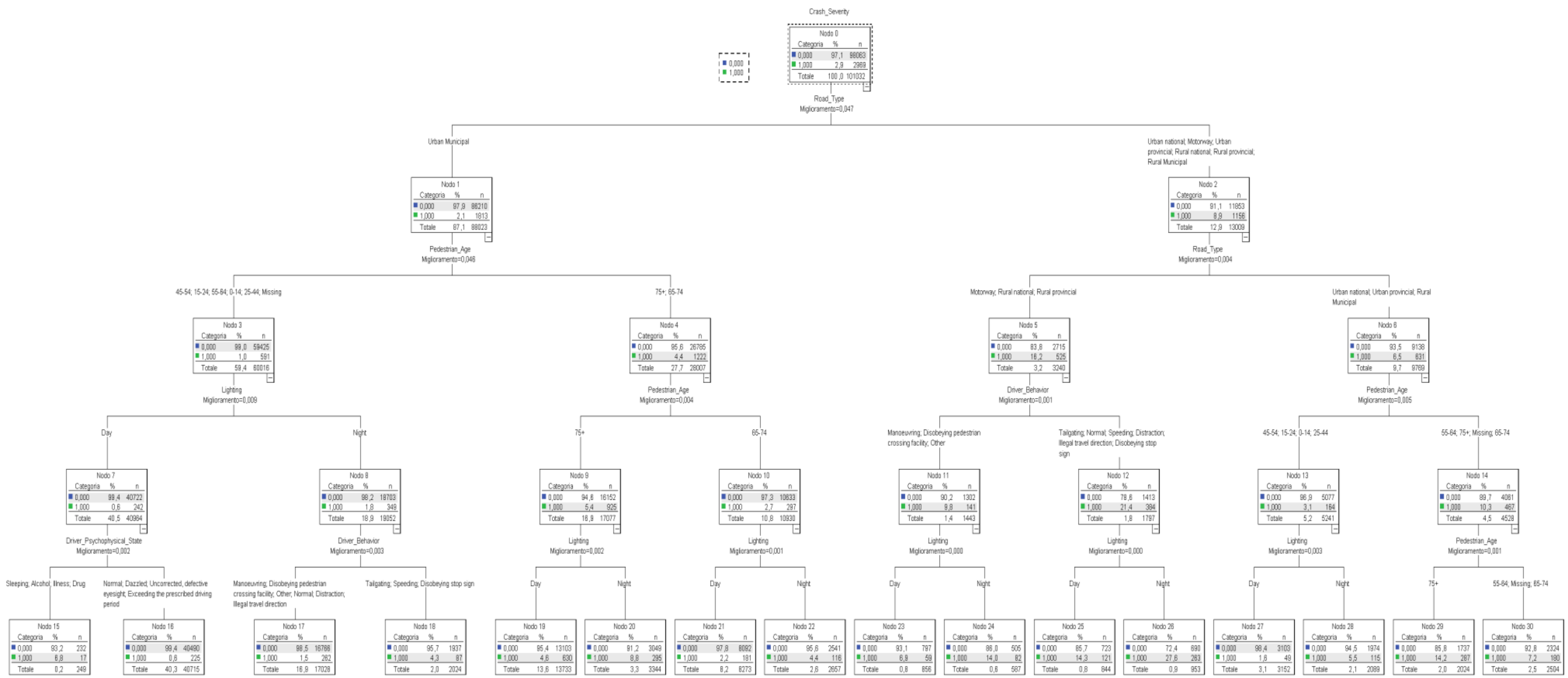


Figure 35 – Classification tree, Italy.

The tool generated 16 terminal nodes, 13 of which predicted fatal crashes and 3 predicted injury crashes (note that Italian database does not provide information related to slight and serious crashes).

The posterior classification ratio (PCR) was assessed for all the nodes (see APPENDIX ITALY, classification tree section) but was reported only for the terminal nodes to understand how representative each terminal node is in relation to the predicted class. The node 19 exhibited the highest PCR (21.22) for fatal classification, followed by nodes 20, 29, and 26 with PCR equal to 9.94, 9.67, and 8.86 respectively.

Table 121 – Terminal nodes and relative Posterior Classification Ratio value, Italy.

Node	PCR		
	Injury	**Fatal**	**Actual Predicted Class**
15	0.24	0.57	Fatal
16	41.29	7.58	Injury
17	17.10	8.82	Injury
18	1.98	2.93	Fatal
19	13.36	21.22	Fatal
20	3.11	9.94	Fatal
21	8.25	6.10	Injury
22	2.59	3.91	Fatal
23	0.81	1.99	Fatal
24	0.51	2.76	Fatal
25	0.74	4.08	Fatal
26	0.70	8.86	Fatal
27	3.16	1.65	Injury
28	2.01	3.87	Fatal
29	1.77	9.67	Fatal
30	2.37	6.06	Fatal

The analysis of variable importance (Figure 36) identified four variables as mostly influencing the classification accuracy of pedestrian crash severity: (1) pedestrian age, (2) road type, (3) area, and (4) lighting.

Table 122 – CT Independent Variable Importance, Italy.

Independent Variable	Importance	Normalized Importance
Pedestrian Age	0.06	100.00%
Road Type	0.05	87.32%
Area	0.04	62.17%
Lighting	0.02	26.65%
Driver Psychophysical State	0.01	16.58%
Vehicle Type	0.01	11.65%
Driver Behaviour	0.01	9.97%
Driver Age	0.00	4.25%
Alignment	0.00	2.92%
Pedestrian Psychophysical State	0.00	2.72%
Pedestrian Gender	0.00	1.78%
Vehicle Defect	0.00	1.50%
Pedestrian Behaviour	0.00	1.41%
Driver Gender	0.00	0.92%
Weather	0.00	0.56%
Vehicle Age	0.00	0.54%
Pavement	0.00	0.49%
Season	0.00	0.10%

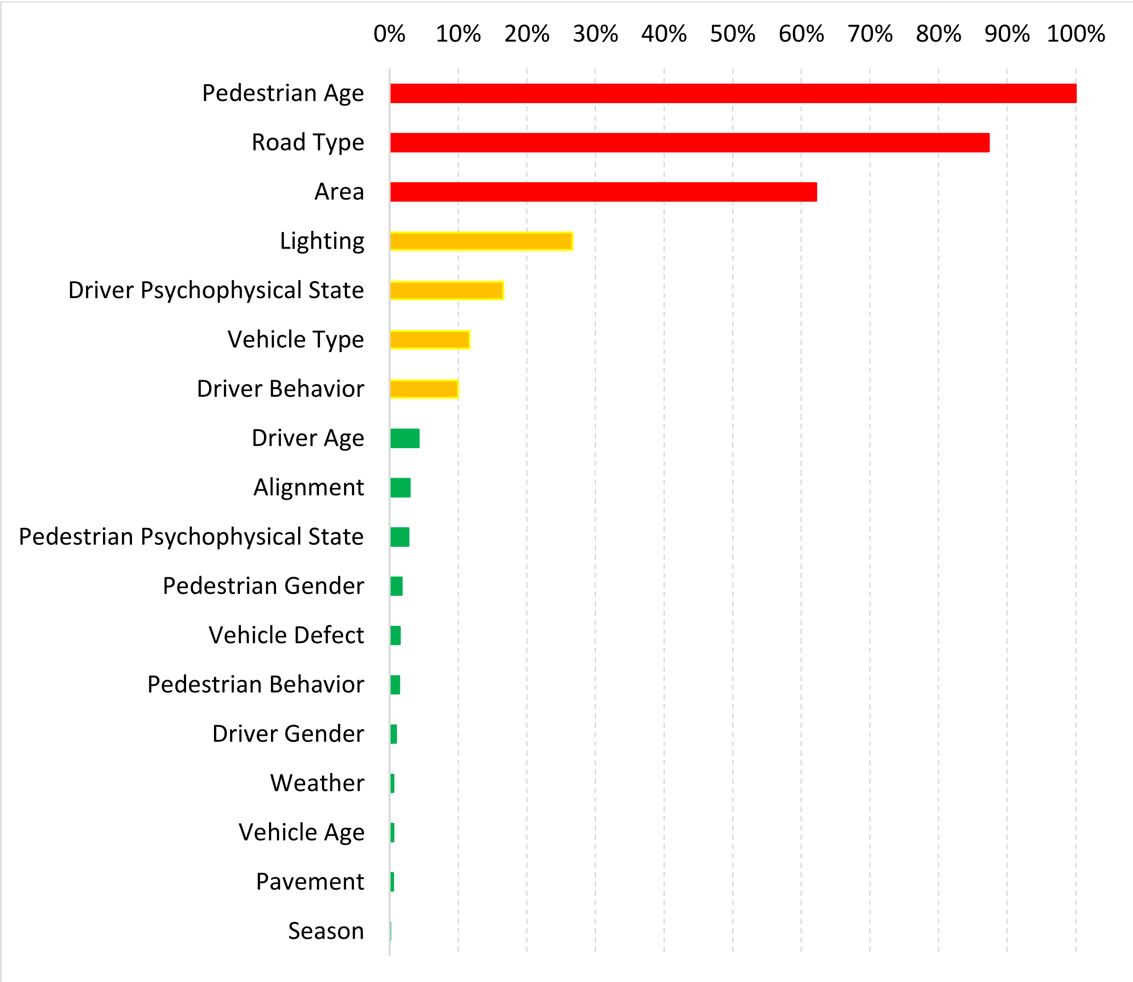


Figure 36 – CT variable normalized importance, Italy.

Overall, the CT tool exhibited 70% for injury, and 76% for fatal crashes with a global accuracy superior to 70%.

Table 123 – Confusion matrix for the Classification tree, Italy.

		Predicted	
		Injury	Fatal
Observed	Injury	68,451	29,612
	Fatal	717	2,252

Table 124 – Performance metrics for the Classification tree, Italy.

Performance metrics	
$TN_{rate}(Acc^{-})$	0.698
TP_{rate} (Acc^{+})	0.759
Precision	0.071
F-Measure	0.129
G-Mean	0.728
AUC	0.788
Acc	0.700
Err	0.300

5.3.4 Random forest

Initially, RF was implemented generating 500 trees. However, after setting the optimal number of trees based on the out of bag (OOB) sample error rate, the RF tool was performed with 40 trees. After determining the number of optimal trees, RF was performed again to determine the list of the most important variables associated with pedestrian crash severity in Italy. For the Italian database, it was only possible to estimate the importance of the predictor variables for fatal crashes. The importance of each explanatory variable showed pedestrian age and alignment contributing most to fatal crashes (Table 125, Figure 38).

All 40 trees were extracted. However, below only the first tree generated by the random forest has been reported. The other trees were reported in the APPENDIX ITALY (random forest section).

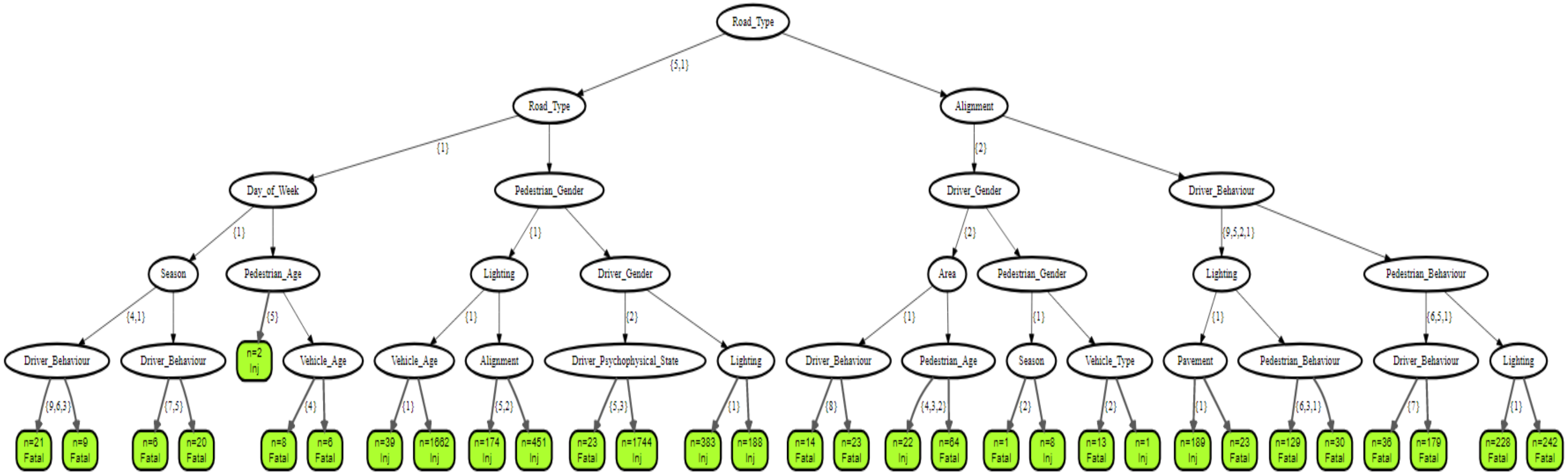


Figure 37 – First tree of the Random Forest, Italy.

Table 125 – RF Independent Variable Importance, Italy.

Independent Variable	Importance	Normalized importance
Pedestrian Age	0.067	100.0%
Alignment	0.037	55.9%
Pedestrian Gender	0.023	34.1%
Driver Gender	0.020	29.9%
Driver Age	0.014	21.5%
Weather	0.004	5.4%
Driver Behaviour	0.003	4.9%
Vehicle Age	0.003	4.8%
Pavement	0.003	4.7%
Pedestrian Behaviour	0.002	3.5%
Season	0.000	0.2%
Day of Week	0.000	-0.1%
Vehicle Defect	-0.001	-1.2%
Lighting	-0.002	-3.2%
Pedestrian Psychophysical State	-0.004	-6.3%
Involved vehicles	-0.008	-11.7%
Vehicle Type	-0.026	-38.4%
Driver Psychophysical State	-0.095	-142.3%
Road Type	-0.099	-147.5%
Area	-0.121	-181.6%

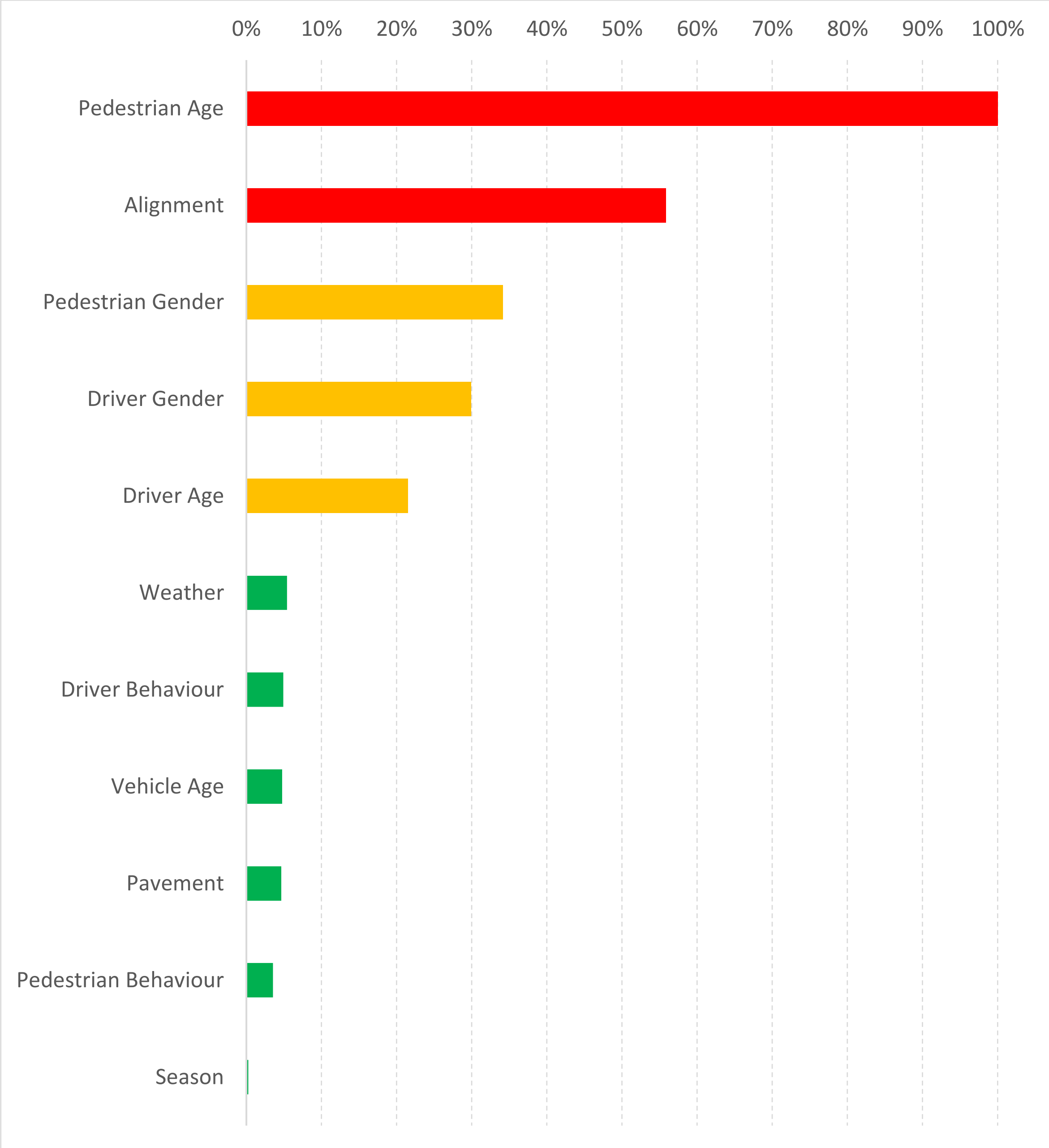


Figure 38 – RF variable normalized importance, Italy.

Overall, the RF tool exhibited 93% for injury and 60% for fatal crashes with a global accuracy equal to 92%. The global accuracy of RF tool was very high and superior to the global accuracy exhibited by CT algorithm.

Table 126 – Confusion matrix for the RF, Italy.

		Predicted	
		Injury	Fatal
Observed	Injury	91,155	6,908
	Fatal	1,183	1,786

Table 127 – Performance metrics for the RF, Italy.

Performance metrics	
$TN_{rate}(Acc^{-})$	0.930
TP_{rate} (Acc^{+})	0.602
Precision	0.205
F-Measure	0.306
G-Mean	0.748
AUC	0.788
Acc	0.920
Err	0.080

5.3.5 Association rules

The threshold values of support (S), confidence (C), and lift (L) were set as follows: S ≥ 0.1%, C ≥ 3%, L ≥ 1.5, and LIC ≥ 1.05.

For pedestrian crash severity were 194 the rules satisfying the predefined thresholds in terms of support, confidence, lift, and lift increase. 23 rules included roadway factors as first antecedents, 135 rules were generated by pedestrian age older than 75, 6 rules included truck vehicles, 26 rules included driver behaviour, and 4 rules were generated by night-time conditions. The rules extracted were ordered by number of antecedents and by the decreasing value of lift. However, Table 128 contains the strongest rules predicting fatal pedestrian crashes in Italy.

Table 128 – Association rules with fatal as consequent, Italy.

Antecedents	**S %**	**C %**	**Lift**	**LIC**
Road Type=Rn	17.32	16.53	5.62	n.a.
Road Type=Rn & Lighting=Night	12.47	22.99	7.82	1.39
Area=Rural	63.94	13.24	4.51	n.a.
Area=Rural & Driver behaviour=Speed	13.96	23.46	7.98	1.77
Area=Rural & Pedestrian Age≥75	13.36	20.61	7.01	1.56
Area=Rural & Pedestrian Age≥75 & Vehicle Type=Car	11.09	21.71	7.39	1.05
Area=Rural & Lighting=Night	40.68	19.98	6.80	1.51
Area=Rural & Season=Winter	17.22	14.09	4.79	1.06
Road Type=Urban provincial	33.26	6.76	2.30	n.a.
Road Type=Urban provincial & Pedestrian Age≥75	16.83	14.82	5.04	2.19
Pedestrian Age≥75	1.30	6.75	2.30	n.a.
Pedestrian Age≥75 & Driver behaviour=Speed	19.70	12.05	4.10	1.79
Pedestrian Age≥75 & Driver behaviour=Speed & Pedestrian Gender=Male	12.77	15.54	5.29	1.29
Pedestrian Age≥75 & Driver behaviour=Speed & Pedestrian Gender=Male & Driver Gender=Male	10.89	17.32	5.89	1.11
Pedestrian Age≥75 & Lighting=Night	46.92	11.73	3.99	1.74
Pedestrian Age≥75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	17.72	16.20	5.51	1.38
Pedestrian Age≥75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Vehicle Age=0-10	10.29	17.72	6.03	1.09
Pedestrian Age≥75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Pavement=Dry	12.08	17.23	5.86	1.06
Pedestrian Age≥75 & Lighting=Night & Driver behaviour=Normal & Alignment=Tangent	11.78	17.42	5.93	1.18
Pedestrian Age≥75 & Lighting=Night & Pedestrian Gender=Male & Driver Age=25-44	10.49	17.04	5.80	1.19
Pedestrian Age≥75 & Lighting=Night & Alignment=Tangent & Vehicle Age=0-10	19.80	15.89	5.41	1.14
Pedestrian Age≥75 & Lighting=Night & Alignment=Tangent & Driver Gender=Male	29.50	14.75	5.02	1.06
Pedestrian Age≥75 & Vehicle Type=Truck	17.72	10.37	3.53	1.54
Pedestrian Age≥75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	41.67	8.97	3.05	1.33
Pedestrian Age≥75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Pedestrian Gender=Male	26.03	11.40	3.88	1.27
Pedestrian Age≥75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Day of Week=Weekend	10.19	10.60	3.61	1.18
Pedestrian Age≥75 & Driver Age=18-24	11.28	8.32	2.83	1.23
Pedestrian Age≥75 & Day of Week=Weekend	31.48	7.82	2.67	1.16
Pedestrian Age≥75 & Day of Week=Weekend & Alignment=Tangent	22.86	9.45	3.22	1.21
Pedestrian Age≥75 & Alignment=Tangent & Pavement=Wet	15.34	8.56	2.91	1.11
Vehicle Type=Truck	41.37	5.97	2.03	n.a.
Vehicle Type=Truck & Lighting=Night	11.09	8.66	2.95	1.45
Driver behaviour=Speed	53.94	5.79	1.97	n.a.
Driver behaviour=Speed & Lighting=Night	30.49	9.83	3.35	1.70
Driver behaviour=Speed & Lighting=Night & Vehicle Age=0-10	15.84	11.89	4.05	1.21

Overall, the AR tool exhibited 42% of correct classification for injury crashes and 91% for fatal crashes with a global accuracy superior to 43%.

Table 129 – Confusion matrix for the AR, Italy.

		Predicted	
		Injury	Fatal
Observed	Injury	41,175	56,888
	Fatal	280	2,689

Table 130 – Performance metrics for the AR, Italy.

Performance metrics	
$TN_{rate}(Acc^{-})$	0.420
TP_{rate} (Acc^{+})	0.906
Precision	0.045
F-Measure	0.086
G-Mean	0.617
AUC	0.774
Acc	0.566
Err	0.434

5.3.6 Support vector machine

SVM model was performed with RBF kernel function. The model returned 46,619 support vectors defining the complex hyperplane. The importance of variables showed the predictors mostly contributing to the classification process. Table 131 and Figure 39 resumed the predictors mostly contributing to the correct classification of pedestrian crash severity identified by SVM: driver psychophysical state, pedestrian psychophysical state, vehicle defect, and road type.

Table 131 – SVM Independent Variable Importance, Italy.

Independent variables	**Importance**	**Normalized Importance**
Driver Psychophysical State	0.268	100.0%
Pedestrian Psychophysical State	0.246	92.1%
Vehicle Defect	0.115	43.0%
Road Type	0.109	40.6%
Weather	0.060	22.5%
Alignment	0.055	20.4%
Driver Behaviour	0.052	19.6%
Pedestrian Age	0.017	6.3%
Pavement	0.016	6.1%
Driver Age	0.016	5.9%
Vehicle Type	0.011	4.1%
Pedestrian Behaviour	0.010	3.9%
Driver Gender	0.005	2.0%
Vehicle Age	0.005	1.9%
Area	0.004	1.4%
Season	0.003	1.0%
Lighting	0.002	0.8%
Day of Week	0.001	0.3%
Pedestrian Gender	0.001	0.3%

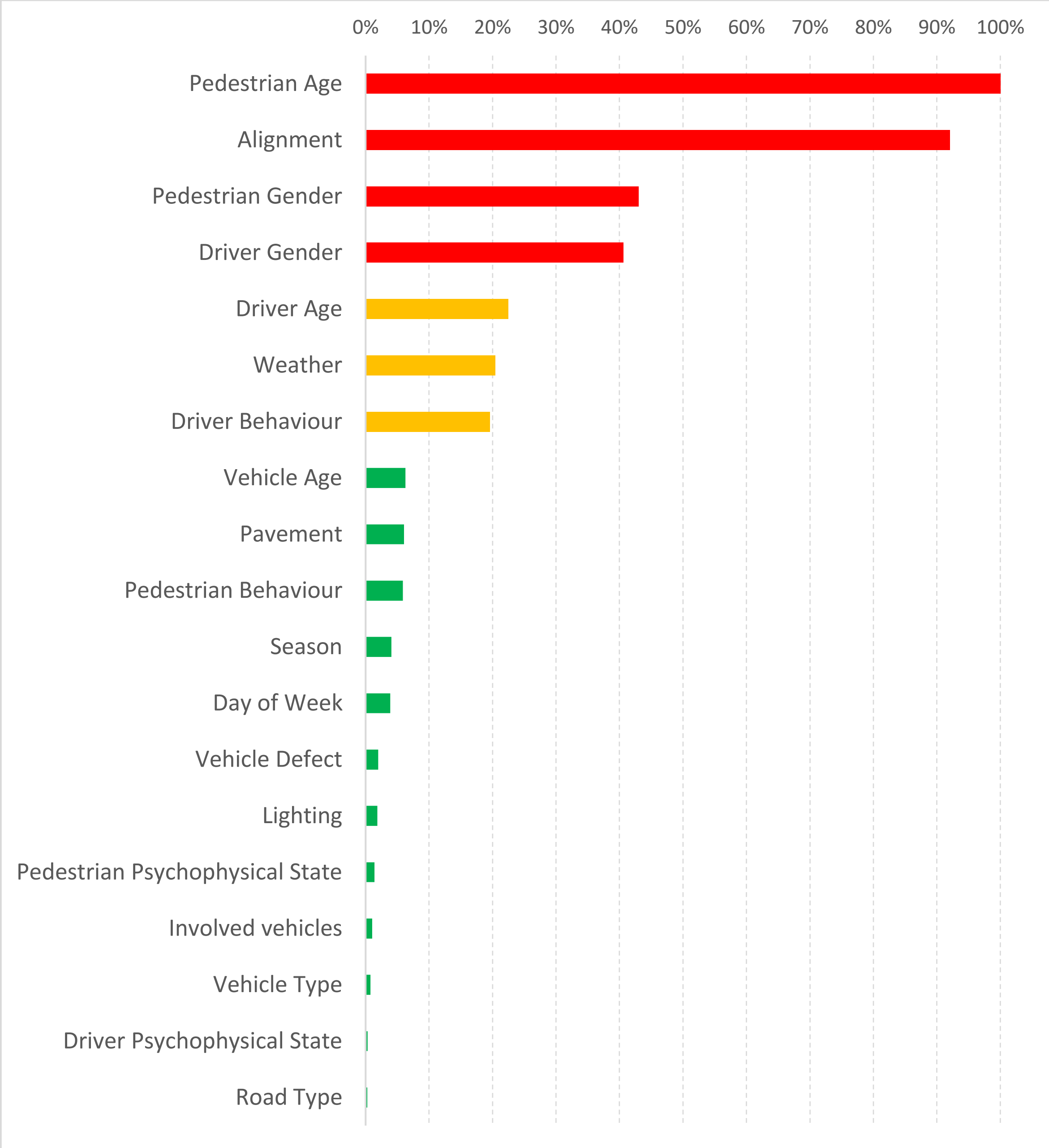


Figure 39 – SVM variable importance, Italy.

Overall, the SVM tool exhibited 99% of correct classification for injury crashes and 61% for fatal crashes with a global accuracy equal to 98%. Over the British and Swedish case studies, SVM reached the highest classification accuracy also in the Italian context.

Table 132 – Confusion matrix for the SVM, Italy.

		Predicted	
		Injury	Fatal
Observed	Injury	97,298	765
	Fatal	1,156	1,813

Table 133 – Performance metrics for the SVM, Italy.

Performance metrics	
TN_{rate}(Acc-)	0.992
TP_{rate} (Acc+)	0.611
Precision	0.703
F-Measure	0.654
G-Mean	0.778
AUC	0.695
Acc	0.981
Err	0.019

5.3.7 Artificial neural network

The ANN tool provides a network made up by 20 factors (excluding the bias) namely day of week, season, lighting, road type, area, alignment, pavement, weather, involved vehicles, vehicle type, vehicle age, driver behaviour, vehicle defect, driver psychophysical state, driver age, driver gender, pedestrian behaviour, pedestrian psychophysical state, pedestrian gender, and pedestrian age.

Table 134 – Artificial Neural Network general information, Italy.

Network Information			
Input Layer	Factors	1	Day of Week
		2	Season
		3	Lighting
		4	Road Type
		5	Area
		6	Alignment
		7	Pavement
		8	Weather
		9	Involved vehicles
		10	Vehicle Type
		11	Vehicle Age
		12	Driver Behaviour
		13	Vehicle Defect
		14	Driver Psychophysical State
		15	Driver Age
		16	Driver Gender
		17	Pedestrian Behaviour
		18	Pedestrian Psychophysical State
		19	Pedestrian Gender
		20	Pedestrian Age
	Number of Units[a]		97
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		14
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Crash Severity
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

Overall, 97 predictor indicators were connected with 14 units belonging to the hidden layer through the hyperbolic tangent activation function. Then, through the softmax function, the input layer predictors were finally linked with the output layer represented by the crash severity variables which, based on Italian data, in a binary structure. The parameter estimates provided by the tool were reported in the appendix (the reader refers to APPENDIX ITALY, Artificial Neural Network section)

Table 135 – ANN Independent Variable Importance, Italy.

Independent Variable	Importance	Normalized Importance
Pedestrian Age	0.136	100.0%
Driver Behaviour	0.085	62.6%
Road Type	0.076	55.7%
Vehicle Type	0.066	48.5%
Area	0.065	47.9%
Driver Age	0.063	46.2%
Lighting	0.063	46.0%
Driver Psychophysical State	0.060	44.0%
Alignment	0.059	43.6%
Pedestrian Behaviour	0.042	30.8%
Driver Gender	0.038	28.0%
Vehicle Age	0.037	27.4%
Weather	0.037	27.1%
Season	0.028	20.8%
Pedestrian Gender	0.028	20.8%
Pedestrian Psychophysical State	0.027	20.2%
Pavement	0.027	19.8%
Involved vehicles	0.027	19.7%
Day of Week	0.020	14.6%
Vehicle Defect	0.016	11.5%

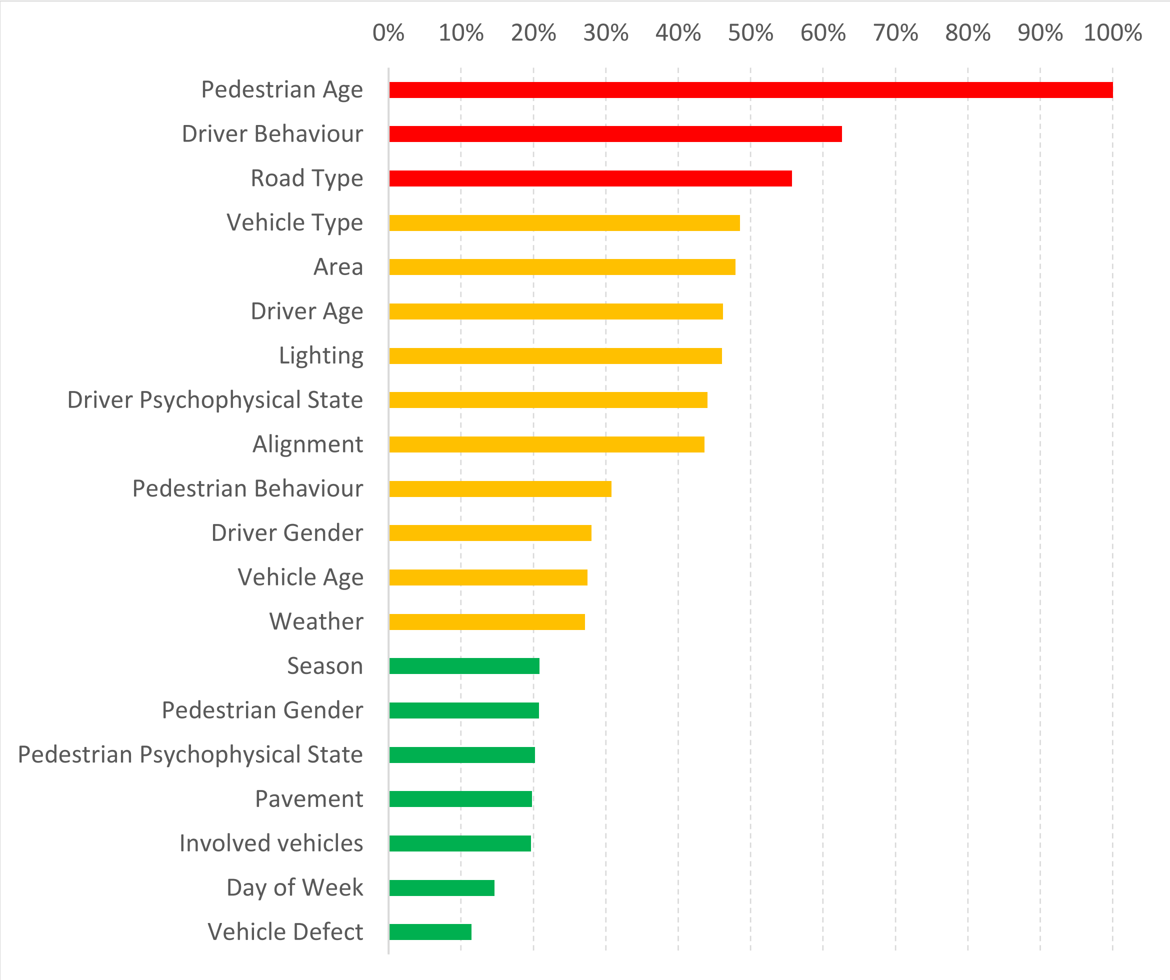


Figure 40 – ANN variable importance, Italy.

Overall, the ANN tool exhibited 75% of correct classification for injury crashes and 75% for fatal crashes with a global accuracy equal to 75%.

Table 136 – Confusion matrix for the ANN, Italy.

		Predicted	
		Injury	Fatal
Observed	Injury	73,258	24,805
	Fatal	729	2,240

Table 137 – Performance metrics for the ANN, Italy.

Performance metrics	
TN_{rate}(Acc-)	0.747
TP_{rate} (Acc+)	0.754
Precision	0.083
F-Measure	0.149
G-Mean	0.751
AUC	0.825
Acc	0.747
Err	0.253

5.3.8 Synthesis of the results

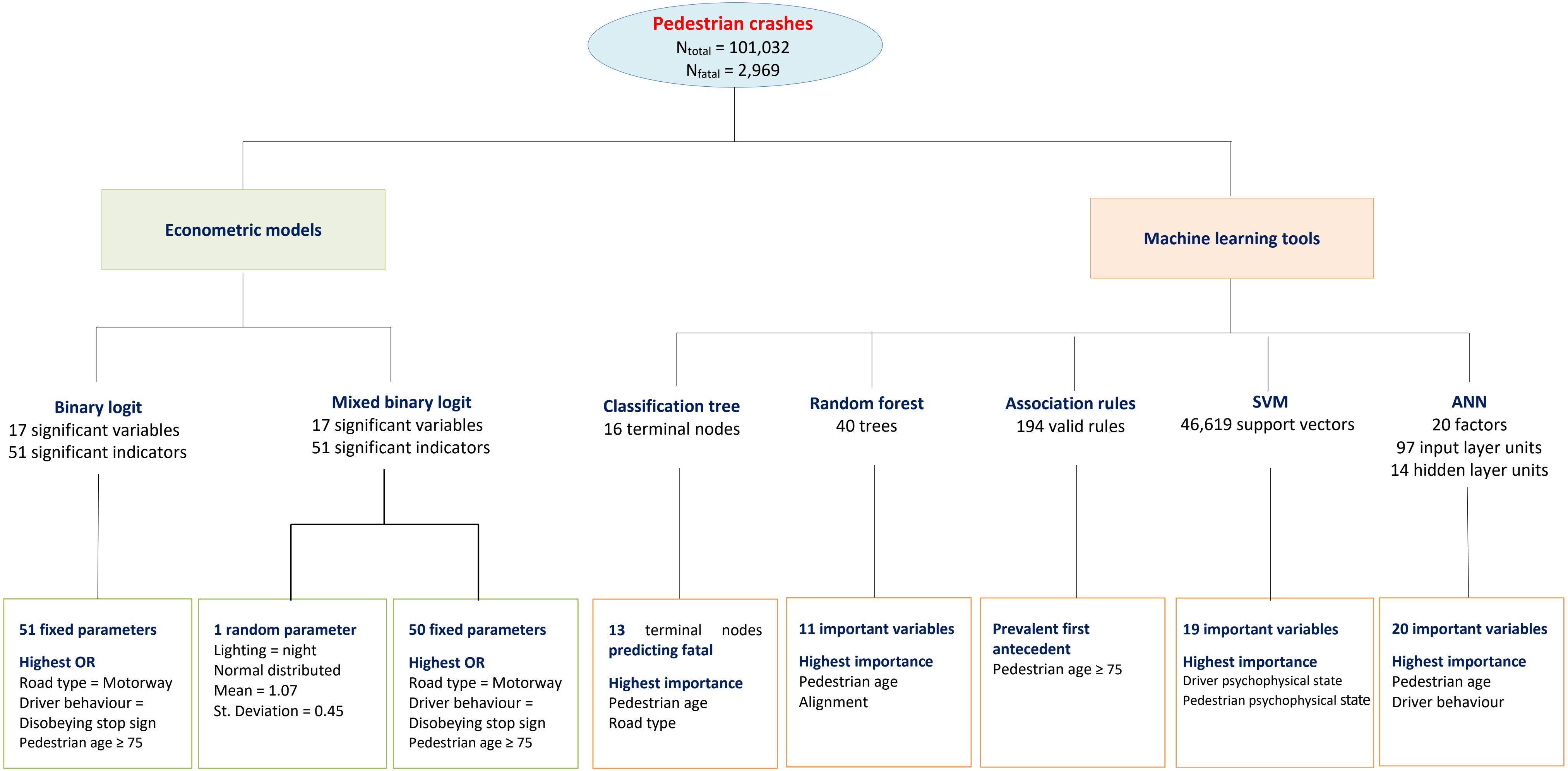


Figure 41 – Main results of the econometric models and machine learning tools, Italy.

5.3.9 Measures of performance

As for the British case study, models' performances were evaluated by the F-measure, G-mean, and AUC. Given that the implementation of the cost-sensitive approach improves all the methods classification performances (as demonstrated for the British dataset), the measures of performance exhibited by the models were provided considering the weighted formulation. Results of the methods are shown in Table 138. The results were provided only for fatal crashes as the Italian database does collect information regarding the crash severity by distinguishing only two levels of severity (injury and fatal crashes) without providing insights on entity of the injuries (slight injury or serious injury). As a consequence of a binary response variable, the ordered models (Ordered Logit – OL – and Random Parameter Ordered Logit – RPOL) cannot be implemented.

Table 138 – Measures of performance of weighted econometric models and machine learning algorithms, Italy.

	Econometric models			**Machine learning**				
	Logit	**RPLogit**		**AR**	**CT**	**RF**	**ANN**	**SVM**
				Fatal				
F-measure	0.15	0.16		0.09	0.13	0.31	0.15	0.65
G-mean	0.76	0.76		0.62	0.73	0.75	0.75	0.78
AUC	0.84	0.84		0.77	0.79	0.79	0.83	0.70

The performances of the methods applied to the Italian database reveal that the logit models (both the fixed and the random formulations) provide similar predictive performances. The value assessed for each metric differ very slightly and it can be observed rounding off to 3 digits after the decimal point. However, the RPLogit provides additional insights on the distribution of parameters in the empirical analysis by the identification of the random distributed variable indicator night-time.

As far as the machine learning tools are concerned, SVM outperformed the other algorithms in classifying fatal crashes reaching accuracy in both correct positive and negative case classification equal to 78% and an F-measure value equal to 0.65, meaning that the model's good overall performance assessed considering also the minority class is equal to 65%. Among the methods implemented in the Italian case study, this is the highest F-measure value observed, followed by RF with an F-measure value equal to 0.31. However, ANN exhibited relevant G-mean and AUC values. G-mean, for instance, is second only to G-mean for SVM.

Overall, SVM and RF have the best predictive performances providing high values for each metrics assessed.

5.3.10 Significant explanatory variables and effects on crash severity

In Table 139, the significant explanatory variables associated with an increase in crash severity were summarized. Table 139 contains variables associated with an increase in fatal crash probability. 16 variables are significant both in the econometric models as well as in the machine learning algorithms and 3 variables are significant only in the machine learning algorithms. The variables significant only in the machine learning algorithms are area, number of vehicles involved, and vehicle age.

Table 139 – Variables associated with an increase in fatal crash probability, Italy.

Both econometric/ML models	Econometric models	ML algorithms
Alignment	-	Area
Day of Week		Involved vehicles
Driver Age		Vehicle Age
Driver Behaviour		
Driver Gender		
Driver Psychophysical State		
Lighting		
Pavement		
Pedestrian Age		
Pedestrian Behaviour		
Pedestrian Gender		
Pedestrian Psychophysical State		
Road type		
Season		
Vehicle Defect		
Vehicle Type		
Weather		

5.3.9.1 Roadway characteristics

Rural area was highlighted as patterns influencing pedestrian crash severity by CT and AR. In particular, the strong association between rural area and the most serious crashes results in a very significant two-item rule with a lift equal to 4.51 (rule 5). Moreover, the variable area was among the most significant predictors among almost all machine learning methods. However, for CT and ANN area showed 62.2% and 47.9% importance in classification respectively.

Road type and alignment were identified as contributory factors by both two groups of methods. Considering urban municipal as the base condition, the econometric models showed that all road types were connected to significant positive estimates indicating a greater propensity to fatal pedestrian crashes on these roads characterized by higher operating speeds and reduced, sometimes unexpected, presence of pedestrians rather than on the urban municipal roads. AR confirmed the result provided by the econometric models whereas, even though road type was the first split for CT growth, the terminal node of the tree with the highest classification accuracy (node 19) belong to the urban municipal branch. Regarding the alignment, the logit, the mixed logit, and AR highlighted the tangents as contributory factors for pedestrian fatality. Specifically, tangent was strongly associated with elder pedestrians (aged over 75). In this case, the presence of tangent alignment increases the probability of a fatal crash by 15% (rule 100). By contrast, intersections (both signalized or unsignalized) and roundabouts were associated with a decrease in crash severity. Alignment was also the variable whose importance in classification was second only to pedestrian age in the RF tool whereas the presence of the variable alignment exhibited 43.6% normalized importance in ANN.

5.3.9.3 Vehicle characteristics

The type of vehicle involved in a pedestrian crash has a significant impact on crash severity. Specifically, a pedestrian struck by a truck has the higher attendance risk. The results were highlighted by all methods. By contrast, a pedestrian crash with the involvement of PTWs has less probability of being fatal. Furthermore, all the machine learning tools found a correlation between vehicles age and fatal pedestrian crashes. New vehicles (vehicles registered less than 10 years ago) were found to have a positive effect on crash severity contributing to fatal pedestrian crashes. Even though both the econometric models did not provide statistically significant estimates for the variable vehicle age, both models identified that possible vehicles defects have a significant effect on fatal pedestrian crashes.

5.3.9.2 Environmental characteristics

The day of the week was identified both by the logit and mixed logit models, AR, and ANN. However, the result of econometric models suggested that, in comparison with the weekday, it is during the weekend that the severity of crashes dramatically increases and this result is also consistent with association rule findings. For the season indicator variable, autumn and winter were significant both in mixed logit and association rules. Night-time involves an increase in pedestrian fatality. The factor night-time generated a two-item rule with a lift equal to 1.63 (rule 191). The association of night-time with the summer (by almost 40%) and the spring (by almost 25%) increases the probability of fatal pedestrian crashes.

Wet pavement was identified as critical by association rules influencing pedestrian outcomes. Slippery pavement was also responsible for fatal crashes. Raining weather condition was identified as a factor contributing to the most severe pedestrian crashes only by association rules.

5.3.9.4 Crash characteristics

The number of vehicles involved in the crash has almost 20% of importance in the ANN classification process.

5.3.9.5 Driver characteristics

The driver behaviour variable was significant for all models, even if the indicator variables exhibited different effects depending on the group of the implemented model. The logit and the mixed logit estimated the significant effect of drivers adopting inappropriate behaviour (such as disobeying stop sign, distraction, speeding, and tailgating) increasing the severity of pedestrian crashes. Driver behaviour was also the split that gave rise to node 26 (one of the strongest nodes of the tree) in CT predicting fatal pedestrian crashes in presence of drivers adopting unsafe behaviours, especially during the night-time. The AR confirmed the speeding and further identified drivers disobeying pedestrian crossing facilities. Driver behaviour was also the second most important variable in the ANN classification process with normalized importance equal to 62.6%.

The relation between altered driver psychophysical state and fatal pedestrian crashes was identified by the econometric models with driving exceeding the prescribed driving period, illness, driving under drug influence, and sleeping having the highest estimates. The driver psychophysical state variable appeared among the most important predictors in CT and ANN. Specifically, the variable generated one of the last split of the tree in CT, with illness, driving under drug or alcohol influence, and sleeping contributing to fatal crashes with pedestrian involvement.

The correlation between driver age and pedestrian crashes was identified by both the logit (fixed/mixed) models and by RF, AR, and ANN. As far as the econometric models are concerned, young drivers (18-24 years old) exhibited positive estimates meaning an increase in the likelihood of fatal crashes when they are involved in a pedestrian crash. The variable was further among the sixth for importance in ANN classification with 46.2% of importance and the fifth in the RF learning process with 21.5% of importance. The rule discovery tool found very young drivers more likely to be involved in fatal pedestrian crashes with elder pedestrians (at least 75 years old) generating the strongest rule among those associating pedestrian age ≥ 75 and driver age (rule 83, lift = 2.83, LIC

= 1.23). For driver gender, the econometric methods found male driver involvement in the most serious pedestrian crashes. Result confirmed by the AR tool.

5.3.9.6 Pedestrian characteristics

Pedestrian characteristics were further identified as potential contributory factors to fatal pedestrian crashes. Pedestrians crossing outside pedestrian crossing facilities are more prone to fatal crashes. The result was highlighted by both groups of methods, with the mixed logit further identifying a decrease in fatal pedestrian crashes when pedestrians cross using the dedicated facilities. Alcohol was the unique pedestrian psychophysical state indicator which results significant in the econometric models. A pedestrian psychophysical state altered by alcohol abuse leads to an increase in fatal pedestrian crash likelihood. The variable is the second strongest split in SVM with normalized importance equal to 92.1%. With less importance, albeit significant, the variable appeared as important also in the ANN process with 20.2% of normalized importance.

As far as pedestrian age is concerned, AR found elder drivers (at least 75 years old) more likely to be involved in fatal pedestrian crashes. The result was consistent with those provided by econometric models. The fixed/mixed logit models further identified a gradually increase in the probability of fatal pedestrian crashes with pedestrian age. Lastly, several rules revealed that male pedestrians were significantly associated with fatal crashes.

CHAPTER VI ~ DISCUSSION

This chapter provides a discussion of the main results obtained in this research. Results were pointed out in different sections according to: consideration on the results provided by econometric models only, consideration on the results provided by machine learning tools only, considerations on all model performances in presence of different sample sizes, and then the main contributory factors identified were discussed and possible safety countermeasures were proposed. In their relative sections, the pros and cons of each group of methods were identified and discussed.

6.1 Considerations on econometric models

Downstream the implementation of the econometric models and the machine learning tools to the three different National databases, some general conclusions can be drawn.

Regarding the econometric models applied to the British context, the results showed that the unordered model (both fixed and random-parameters multinomial logit models) exhibited acceptable predictive performance and a superior fit to the ordered models (both fixed and random-parameters ordered logit models). The ordered models, indeed, showed poor ability in classifying correctly fatal crashes even after the implementation of the weighting procedure with the lowest performances achieved by the standard OL. The results of this thesis are also consistent with previous studies (Abay, 2013; Eluru et al., 2008; Savolainen et al., 2011).

On the other hand, from the application of the models to the Swedish context, the OL exhibited a superior fit to MNL and RPMNL. However, the model identifies fewer significant variables and indicators. One of the main explanations for this difference is that the ordered probability models place a strict restriction on how exogenous variables affect outcome probabilities. This implies that the OL does not allow the probabilities of both the highest and lowest severity levels to increase or decrease. Thus, to an increase in the probability of the highest severity class (fatal in this research), a decrease in the probability of lowest severity levels (slight in this research) is observed and vice-versa. Thus, this research confirms that implementing the ordered crash severity nature on logistic regression models does not necessarily improve the predictive performances across all severe levels meaning that the relationships between predictors and crash severity outcomes might not be monotonous. Accordingly, simply because the values of a variable can be ordered does not imply that the variable should be analysed as ordinal (Long, 1997).

The use of the random parameter models provided evidence of the existence of heterogeneity among data. All the significant variables impacting pedestrian crash severity in the standard logit models were tested for heterogeneity. Different indicator variables showed normally distributed random parameters, with statistically significant standard deviations, in the different case studies:

In the British case study: the RPMNL identified two random variables for fatal crashes (going ahead as vehicle manoeuvres and roundabout) and one for serious crashes (pedestrian age greater or equal to 75) whereas the RPOL found one random variable for both severity levels (pedestrian age greater or equal to 75). In the Swedish case study: the RPMNL identified one random variable for fatal crashes (speed limit ≤ 30 km/h) and one for serious crashes (roundabout). In the Italian case study, the RPLogit identified one random variable for fatal crashes (night-time).

The presence of such variability in the effect of variables across the sample population highlights the need to account for potential unobserved heterogeneity across vehicle-pedestrian crashes that may improve understanding and reduce erroneous inferences and predictions, producing more accurate and informative results. This is also a possible explanation to the better classification performances achieved by the random parameter models compared with their standard formulations.

In line with the above, it is the RPMNL the model with the best predictive performances.

A drawback of the econometric models is that they tend to identify a smaller set of significant variables than machine learning tools. Among the econometrics, the unordered models uncover more insights than the ordered models.

6.2 Considerations on machine learning algorithms

The application of machine learning tools (also known data-driven methods) is a relatively recent practice driven by the need to overcome the issues of traditional statistical modeling.

Among the five implemented machine learning tools, SVM outperformed the other algorithms followed in some cases by RF and in others by ANN. Despite their straightforward capability in achieving very high performances, some machine learning tools act as a black box as the nature of these tools cannot directly correlate crash contributory factors and crash severity nor can they be used to quantify the impact of the contributing factors on injury severity probabilities. Some algorithms, such as CT and RF can be graphically displayed as a tree and their structure enhances comprehension with understandable results. A common output of machine learning models is the importance of variables during the classification process which provides a rank of the explanatory variables in terms of their significance. Unlike most machine learning approaches, AR identifies specific patterns associated with pedestrian crashes, giving strength to the co-occurrence of several factors affecting crash severity. For instance, contributory factors associated with pedestrian crashes are the patterns with higher lift values which can be considered as the parameter for determining the significance of the pattern from the base condition (Das et al., 2019). Furthermore, the rule structure allows a clear framework of attributes combinations.

In addition, the machine learning algorithms provide results that agree on several aspects with the econometric outputs. For instance, the variable importance lists provided by the machine learning tools were consistent with the significant variables identified by the econometric models. Moreover, AR tool often provides associations of independent variables with crash severity whose direction confirm the econometric models' results.

6.3 Consideration on model performances and different sample sizes

The econometric models and the machine learning tools were implemented in three different National databases of different size to evaluate how the information reported in different data and data sample sizes can impact on the results provided by the model performances and results.

Table 140 – Comparison of the MNL and RPMNL results in the three case studies.

	MNL/Logit			**RPMNL/RPLogit**			**OL**	
	Great Britain	Sweden	Italy	Great Britain	Sweden	Italy	Great Britain	Sweden
n. crashes	67,356	9,426	101,032	67,356	9,426	101,032	67,356	9,426
Fatal								
F-measure	0.28	0.15	0.15	0.53	0.14	0.16	0.00	0.20
G-mean	0.50	0.64	0.76	0.65	0.81	0.76	0.04	0.75
AUC	0.87	0.85	0.84	0.94	0.86	0.84	0.85	0.89
Serious								
F-measure	0.21	0.10		0.41	0.23		0.41	0.10
G-mean	0.36	0.47	na	0.58	0.63	na	0.43	0.50
AUC	0.62	0.60		0.68	0.49		0.61	0.59

Table 141 – Comparison of the AR, ANN, and SVM results in the three case studies.

	AR			**ANN**			**SVM**		
	Great Britain	Sweden	Italy	Great Britain	Sweden	Italy	Great Britain	Sweden	Italy
n. crashes	67,356	9,426	101,032	67,356	9,426	101,032	67,356	9,426	101,032
Fatal									
F-measure	0.05	0.06	0.09	0.18	0.48	0.15	0.95	0.79	0.65
G-mean	0.36	0.18	0.62	0.66	0.82	0.75	0.96	0.86	0.78
AUC	0.79	0.80	0.77	0.78	0.91	0.83	0.88	0.86	0.70
Serious									
F-measure	0.39	0.14		0.26	0.10		0.95	0.73	
G-mean	0.54	0.27	na	0.43	0.53	na	0.96	0.85	na
AUC	0.58	0.39		0.76	0.67		0.76	0.54	

Table 142 – Comparison of the CT and RF results in e three case studies.

	CT			RF		
	Great Britain	Sweden	Italy	Great Britain	Sweden	Italy
n. crashes	67,356	9,426	101,032	67,356	9,426	101,032
Fatal						
F-measure	0.16	0.19	0.13	0.57	0.46	0.21
G-mean	0.72	0.61	0.73	0.77	0.76	0.61
AUC	0.82	0.71	0.79	0.88	0.95	0.79
Serious						
F-measure	0.29	0.14		0.90	0.17	
G-mean	0.46	0.62	na	0.92	0.58	na
AUC	0.47	0.63		0.71	0.79	

The performances exhibited by each method were compared in relation to the method itself and the number of crashes in the databases. Even though there is not a clear trend explaining the performances of the models in relation to the data sample size, some considerations can be drawn.

In Table 140 the performance of MNL, RPMNL and OL were reported. The comparison cannot be provided for RPOL as results were obtained only for the British database. OL comparison is provided only for British and Swedish databases as the Italian database structure, the variable crash severity is binary and this does not allow to consider an ordinal nature of severity levels. Unordered models, both fixed and mixed, provide the highest performances when applied to British and Swedish data. It may imply that up to a sample size of the same magnitude of British data, the econometric models still perform well.

Table 140 and Table 141 resume the results obtained by the machine learning algorithms. The results revealed that AR performs better in presence of a wide range of data and a considerable set of variables as for the case of the Italian and the British databases. A similar response was observed by SVM, especially with Great Britain data where all values of the performance metrics for SVM pointed out the maximum classification accuracy of the algorithm. Unexpectedly, ANN performed better with the Swedish data (with less than 10,000 observations), whereas the ANN performance in Great Britain (roughly 70,000 observations) and Italy (roughly 100,000 observations) were almost comparable. As far as CT is concerned, the algorithm did not exhibit the highest performances in relation to a specific national database meaning that the performances of the model are not particularly sensitive to the sample size. However, in presence of small sample sizes (i.e., the Swedish case study), the set of significant variables important to the classification process is very poor. By contrast, RF performed better with the British data, both for fatal and serious crashes.

Overall, the results exhibited by the machine learning methods, made an exception for ANN, may suggest that in presence of small data sizes tend to decrease the advantages of these algorithms to uncover causality. On the other hand, the machine learning methods are also known as data-driven tools which implies that the learning process is guided by the data so that it is already expected to observe better performances in presence of big-data.

It is also noteworthy to mention that the use of the weighted approach for imbalanced data may have enhanced the models' performances as well.

6.4 Main factors identified and safety countermeasures

Using the three different databases, several factors significantly increasing the probability of fatal and serious injury in pedestrian-vehicle crashes were found. Some factors were significant in all case studies meaning that their critical presence is a generalized issue. Among them, darkness, rural area, truck involvement, and pedestrians aged over 75.

As expected, pedestrian crashes occurred during the night or under low-light conditions increased the likelihood of fatal consequences (Noh et al., 2019). The driver may fail to see a pedestrian at night (i.e., in the British case study, night-time fatal pedestrian crashes were also associated with frontal vehicle impact). This pattern highlights the importance of improving pedestrian conspicuity. Babić et al. (2021) found that drivers showed more active eye movements after noticing pedestrians in reflective vests than after noticing pedestrians in non-reflective clothing. Other than reflective clothes and markings, some studies (Fekety et al., 2016; Wood et al., 2017) examined elements of clothing (electroluminescent panels) that may be a useful supplement since they are visible even when a pedestrian is not illuminated by approaching headlamps. Nevertheless, roads should be effectively illuminated as well, especially in areas where there is a high probability of observing pedestrians such as in the proximity of pedestrian crossings. As a matter of fact, the severity of crashes increases far from intersections. Indeed, the results highlighted that intersections (both signalized and unsignalized) and roundabout are safer than tangents where fatal crashes are more likely to occur, consistently with (Eluru et al., 2008, Demsere-Derry et al., 2010). The most effective solution is to provide lighting with light emitting diodes (LEDs) to improve pedestrian visibility during the nighttime as well as the visibility of other road users for pedestrians. Devices to warn motorists of crossing pedestrians, such as in-pavement warning lights with advance signing, flashing in-curb LEDs and beacons at crosswalks, pedestrian-activated overhead flashing beacons or high-intensity activated crosswalk devices, are effective to increase drivers' attention towards pedestrians and reduce crashes, especially at night (Fitzpatrick et al., 2006; Lantieri et al., 2021).

Rural areas and higher speed limits characterize roads where the most severe crashes occurred. The results obtained for each case study pointed out an upward relationship between vehicular speed and pedestrian fatality. This may be a consequence of a typical rural road configuration with higher vehicle speeds combined with fewer separated facilities for pedestrians, such as sidewalks paths, and trails, compared to urban areas. The issue becomes more critical, especially during the nigh-time.

The type of vehicle involved in a pedestrian crash also influences crash severity. Consistently with previous studies (Aziz et al., 2013; Chen & Fan 2019; Kim et al., 2010; Montella et al., 2011; Tay et al., 2011), the presence of a truck/bus involved in the crash results in a contributory factor affecting higher crash severity due to the larger mass and greater stiffness, the larger area of impact for pedestrians, higher bumper height, blunter geometry, and longer stopping distances compared with other vehicles. Furthermore, the presence of articulated vehicles (or vehicles with trailers) has been identified as contributing to the most severe pedestrian crashes in British and Swedish databases (Italian data does not provide this information). The direct link of fatal/serious crashes and trucks, as well as articulated vehicles, suggests the importance of planning specific routes for trucks. To avoid the transit to heavy vehicles in crucial hot spots such as places highly frequented by pedestrians, it is crucial to establish a road hierarchy giving the highest priority to pedestrians and then to the other road users.

Consistently with previous studies (Chen & Fan, 2019; Kim et al., 2010; Montella et al., 2011), young drivers increased the probability of fatal and serious crashes. A possible explanation is that older drivers tend to drive more carefully and at lower speeds. Hence, as motorists become older, pedestrians are more likely to suffer no injury once in a crash. Male drivers were also more likely to be involved in the most serious crashes and the dissertation results confirm previous findings (Das et al., 2019; Martin & Wu, 2017; Montella et al., 2011; Moral-Garcia et al., 2019). These factors may reflect a typical more aggressive way of driving of young and male drivers. Furthermore, inappropriate driving (including speeding, tailgating, disobeying a stop sign, and distraction), as well as altered driving due to the use of alcohol or drug, were highlighted as contributory factors to the most severe crashes. To reduce pedestrian crashes, programmes are essentially required to enforce existing traffic laws, ordinances for drivers and more stringent speed limits. Furthermore, safety education should be integrated with school programs and safety campaigns should be a government priority task.

Furthermore, elderly pedestrians resulted more prone to severe outcomes relative to younger individuals when in a crash. This is due to an increase in perception and reaction times, their physical vulnerability and fragility, and the suffering of various medical conditions, all of which contribute to their higher injury risk propensity (Chen & Fan, 2019; Eluru et al., 2008; Noh et al., 2019). Low-speed areas may be employed during the weekend to avoid the conflict between motor vehicles and pedestrians. The solution may be especially applied in areas with relevant pedestrian activities, especially for elderly pedestrians.

Poor pavement condition, such as slippery or wet pavement, contributes to fatal pedestrian crashes. This factor was found in all case studies highlighting how low pavement friction increases pedestrian fatality risk. Indeed, a longer braking distance and response time would be required when the pavement surface condition is poor. Increasing the pavement skid resistance can reduce braking time and distance and it is essential, especially in case of emergence manoeuvres.

6.5 A note on the non-availability of the speed limit variable: the case with the Italian database.

It is noteworthy to observe that only in the Italian database the information related to the speed limits is not collected.

Sometimes, the datum can be retrieved from the crash narrative and the crash sketch in the highway police reports. However, the information is lost in the Istat database. This variable is, instead, collected by all other international and national databases (Australia, New Zealand, United States of America, Great Britain, and Sweden).

The speed limit on the road refers to the posted speed limit on a certain highway segment and it is generally the maximum speed limit allowed by the law for road vehicles. It is usually set by national or local government legislatures.

The use of the variable indicating the speed limit has been often used as an explanatory variable in many crash analyses. When the information is not available, some studies have used the highway's design speed even though the vehicle operating speeds may differ substantially on highways with the same design speed. In other cases, the variable related to speeds is not collected at all (as is the case of the Italian database) and speed-related considerations may be drawn on the basis of the road type variable linked with the road area.

The results obtained in the British and Swedish contexts pointed out that the crash severity is highly correlated with the speed limits. In the British case study, the speed limit equal to or greater than 50 mph was highly associated with fatal crashes. In the Swedish case study, the speed limit less than or equal to 30 km/h was found to be random. For 76.9% of the observations the presence of speed limits ≤ 30 km/h decreases the probability of fatal crashes while for 23.1% of the observations it leads to an increase in that probability. Over 30 km/h, the probability of fatal crashes increases with increasing speed limits.

Certainly, there are several other factors involved in a crash that affects the crash severity but this research demonstrated the strong and significant increase in crash severity associated with the increase in speed limits. The non-availability of the speed limit variable in the Italian database certainly limited the safety countermeasures that can be identified to mitigate crash severity. For instance, among the factors identified by the analysis in the Italian context, there are the rural municipal, rural provincial, and rural national roads correlated with a monotonic increase in fatal crashes. Urban national and urban provincial roads also exhibited higher severity than municipal roads. Indeed, the urban provincial and national roads are segments of rural roads that cross small urban centres and drivers generally maintain high operating speeds. The availability of the variable speed limits may offer a further explanation about the correlation of road type, area, and crash severity. Therefore, using the result of the analyses, transportation planners, decision-makers, and local authorities can find support to plan, design, operate, and manage a safer transportation system to reduce vehicle-pedestrian collisions.

CHAPTER VII ~ CONCLUSIONS

Reducing road crashes with higher injury severity as well as crashes involving vulnerable road users are the most recent challenge in road safety. Reducing them, social costs will also reduce providing a greater impact on the sustainability of transportation systems.

Confirmed by the severity of vehicle-pedestrian crashes, among the road users, pedestrians are the most vulnerable to road potential risks. Hence, further in-depth research is strongly needed to identify which factors, and how, affect crash injury severity. The identification of factors contributing to crash severity is essential to understand the interaction among the road users and the road environment and suggest appropriate safety countermeasures to mitigate the severity of pedestrian crashes. However, the factors contributing to pedestrian severity are usually different than for motor vehicles, so studies should be focused on vehicle-pedestrian crashes to effectively contribute to pedestrian safety improvement. Furthermore, the interpretation of the results provided by the implemented models is essential to plan targeted investments and actions to improve pedestrian safety and road safety worldwide.

Very often an analyst struggles because of the choice of the “perfect” method to implement in safety analysis trying to achieve top prediction accuracy and, at the same time, exhaustive and reliable factors contributing to crash severity. Traditionally, analysts used safety data (crash data made up by roadway, vehicle, environmental, crash, and users’ characteristics) and econometric models (i.e., logit, probit, unordered/ordered models, fixed/random formulations) to carry out crash severity analyses and this often results in accepting a trade-off between prediction accuracy and interpretation. Indeed, a model that predicts well may not necessarily be the best at uncovering contributory patterns. On the other hand, models that are good at uncovering understandable contributory patterns may not be the best for accurate prediction. The existing literature on the crash severity problem, the complexity of the subject, and the several methodological approaches used over time to conduct crash severity analysis has been summarized in chapter 2.

The literature review pointed out that although econometric models still remain the primary method choice to carry out safety analyses, recently, to handle large and complex datasets, machine learning algorithms are paving the way with the aim of uncovering high-dimensional and nonlinear relationships among the data.

To sum up, trying to provide support for the choice of the appropriate model in performing crash severity analyses, this dissertation presents a multi-comparative analysis between the most widely used methods and the most recent machine learning tools applied to the vehicle-pedestrian crash severity analyses. This research aimed at answering the question: “Are these methods eligible for the purpose?”. A description of the econometric models and the machine learning tools used in this research was provided in chapter 3.

The choice of a method over the others depends on the quality of the available data and the sample size. However, the data traditionally used in crash severity analyses are collected by the police, and only rarely, the severity of crashes is assessed using medical reports. This represents only one of the several potential errors in crash data which can also affect analysis results. The issues related to the crash data, how they are collected and the importance of unified information were discussed in chapter 2. A description of the national databases used in this research, their summary statistics and related comments were provided in chapter 4.

Since the main contribution of this research is to answer the question: “Are these methods eligible for the purpose?”, the research findings were provided by a dual perspective (quantitative and quality evaluations). Then, the results were provided in chapter 5 and organized into three sections, each of them related to the specific case study. An overall discussion of the results, the main contributory factors identified and the appropriate countermeasures were provided in chapter 6.

Finally, according to the main objectives of the thesis stated in Table 1, the main conclusions are drawn:

7.1 Investigation of the issue of imbalanced distributions of crash severity levels

To overcome the problem of extremely imbalanced data and the risk of distorted results from the learning process, in this research a weighted approach was used. The effectiveness of this procedure was firstly tested on the British data. Then, ascertained that all models’ performances improved when a specific weight for each severity level was applied, the methods were all directly performed with the weight approach both in the Swedish and Italian case studies. The use of this procedure helped to reduce both false negative and positive cases minimizing errors in the classification process.

7.2 Comparison of the models in predicting pedestrian crash severity by using performance metrics (quantitative evaluation)

The performances of the models were evaluated by the F-measure, the G-mean, and the AUC. The result of this research demonstrated that the machine learning tools actually provided very high performances in prediction accuracy, and some algorithms (SVM, ANN, and RF) also outperformed the econometric models and others algorithms falling under the same umbrella of machine learning tools. For instance, for fatal crash classification, the SVM reached a value of F-measure ranging from 0.65 to 0.95 and a value of G-mean, even superior, ranging from 0.78 to 0.96. These values are considerably greater than the values reached by the best econometric model (the random-parameter multinomial logit) whose performances ranged from 0.16 to 0.53 in terms of F-measure and from 0.65 to 0.81 in terms of G-mean.

However, even though some machine learning tools (SVM and ANN) exhibited very high classification performances, their results are really difficult to interpret. Other machine learning algorithms, instead, such as AR, CT, provided very understandable and intuitive results even if it meant lower prediction accuracy. Among all the implemented models, the RF is the tool that provided considerably high performances and more interesting outputs. The tool, indeed, extracted the variable importance and graphically displayed the tree belonging to the forest.

7.3 Understanding how the ability of a model applied to different databases changes (different in structure, sample size, geographical condition)

The performances of the models were also assessed in relation to different crash databases and data sample sizes. The results revealed that even though there is not a clear trend explaining the phenomenon, some considerations can be drawn. Unordered models, both fixed and mixed formulations, provide the highest performances when applied to the British database which approximately contains 70,000 observations whereas the OL performed better with the Swedish data (Swedish database approximately contains 10,000 observations). It may imply that in presence of small sample sizes the OL may perform reasonably well (even if it produces a small set of significant variables and indicators) and its performances tend to decrease when the number of observations increases as well as the number of potential explanatory variables. On the other hand, the MNL and especially the RPMNL tend to improve their classification accuracy in presence of databases containing up to a number of observations of the same magnitude of the British data.

By contrast, the results of the quantitative evaluation exhibited by the machine learning methods, made an exception for ANN, may suggest that in presence of small data sizes the advantages of these algorithms to uncover causality tend to decrease. For instance, with the Swedish data, the classification tree tool drastically decreases in accuracy and in significant explanatory variables identified, providing, in such cases, very poor results and interpretation of the phenomenon investigated. However, among the non-parametric models, the association rules presented considerable advantages. Rather than confirming a hypothesis and do not require prespecified assumptions, the tool focuses on identifying hidden patterns in the crash data. Besides, it exhibited fair performances even in the case of missing data, thus reducing the tendency to draw biased and inconsistent conclusions and it can handle both small and large datasets, making it suitable for analysing rare events data.

7.4 Comparison of the models by their capability in identifying significant explanatory variables affecting crash severity (qualitative evaluation)

Regarding the qualitative evaluation, the machine learning tools uncovered more hidden correlations among data than the econometric models and provided valuable insights on the interdependence among the several roadway, environmental, vehicle, and road users related factors contributing to the severity of pedestrian crashes. For instance, in the British case study and for fatal crashes, 19 variables were significant both in the econometric models as well as in the machine learning algorithms, 1 variable was significant only in the econometric models and 7 variables were significant only in the machine learning algorithms. In the Swedish case study, 13 variables are significant both in the econometric models as well as in the machine learning algorithms and 5 variables are significant only in the machine learning algorithms. No further variables are identified by the econometric models only. In the Italian case study, 16 variables are significant both in the econometric models as well as in the machine learning algorithms and 3 variables are significant only in the machine learning algorithms.

However, this research pointed out that questionable limitations do also exist for these artificial intelligence techniques. Despite their outstanding performance in prediction, machine learning techniques such as SVM and ANN do not provide easy to interpret outputs. Even though the importance of variables provides a list of the predictors most influencing the classification process, it does not provide information about the directions and magnitude of variable indicators.

Thus, as econometric models, machine learning tools also fall into accepting trade-offs between prediction accuracy and uncovering underlying causality.

7.5 Identification of the interdependences between crash characteristics and fatal pedestrian crashes

Identifying factors that affect crash injury severity and understanding how these factors affect injury severity is critical in planning and implementing highway safety improvement programs. With this aim, contributory factors were investigated for each nation. Although three different databases were used, several factors were found to significantly affect the probability of fatal and serious injury in pedestrian-vehicle crashes whatever the nation is. It is noteworthy to observe that some factors turned out to increase the probability of the most serious crashes in all case studies meaning that their presence is a critical and generalized issue. Particularly, four variables were identified: lighting equal to darkness, area equal to rural, fatal pedestrian crash with truck involvement (truck as vehicle type), and pedestrians aged over 75 as pedestrian characteristics.

These four factors were separately estimated as significant by the econometric models, and they are often associated in different combinations by association rules and classification trees. In this case, the use of a tool that provides easy to interpret outputs and mainly a relation among the factors may help in enhancing the phenomenon understanding. For instance, the joint presence of truck involvement in fatal pedestrian crashes with dark rural roads and elderly pedestrians pointed out the relationship between higher speed roads, with intense traffic of trucks, and pedestrian fatality. This may be a consequence of a typical rural road configuration with higher vehicle speeds combined with fewer separated facilities for pedestrians, such as sidewalks paths, and trails, compared to urban areas. The issue becomes more critical especially during the nigh-time due to the poor visibility of pedestrians and the other road users during the night-time. Moreover, on rural roads, motorists do not expect the presence of pedestrians crossing, so devices to increase drivers' attention towards pedestrians to reduce pedestrian crashes is crucial, especially at night.

7.6 Proposal for improvement of the Italian crash report form

The Italian national database suffers drawbacks and limitations. Thus, this research also proposes improvements in Istat crash report form by considering the chance of their practical implementation.

As expected, the different information stored in the British, Swedish, and Italian databases impacts the results and the magnitude of exploration and awareness of data.

The British data provides a wide range of variables describing the event, the roadway, the environment, the vehicles involved, and the users involved (both drivers and pedestrians). Furthermore, the variable crash severity is collected on three different levels of severity following the EU recommendations. A downside is that data on severity outcomes are based on police judgement only.

The Swedish database stored information collected by the police and hospital reports. If both reports are provided, the crash severity is assessed following the information of the hospital report otherwise the information is derived from police reports. The variable crash severity is collected on four different levels of severity including fatal crashes, serious injury, slight injury, and crashes with property damage only. A drawback of the Swedish database is that data are collected on multiple files and over the difficulty encountered to unify the data, there is much redundant information which implies that, after the pre-processing procedure, the final crash data contains a limited number of variables. Furthermore, no information about the behaviour of the driver or the pedestrian involved in the crash is provided.

Lastly, the Italian database used in this research is made up of a considerable set of variables partially available to all institutes for research and partially available only thanks to the precious memorandum of understanding stipulated between Istat and the University of Naples Federico II. This part of data contains information related to the psychophysical state of both drivers and pedestrians, drivers' and pedestrians' behaviour at the moment of the crash, and the user age provided in a disaggregated form. Despite the richness of the information contained in the set of data, the variable crash severity is still classified as a binary variable (injury and fatal) with a considerable loss of information regarding the entity of injury outcomes. However, the injury classification including slight injury and serious injury is strongly needed for standardization with the other EU crash databases and for the correct implementation of the EU Directive on road safety management (European Commission, 2019d). What is more, in most cases hospital reports are attached to the police reports and accurate assessments of the crash severity might be carried out but the information gets lost in the national data unification process. This represents a non-trivial issue on the data as it does not allow researchers to conduct analysis on both fatal and serious injury levels, it is far from the EU recommendations, and more dramatically, it does not allow to contribute to the sustainable goal belonging to the Agenda ONU 2030 of halving both road victims and serious injuries. Thus, it is important for the National Institute of Statistics to modify the way crash severity is included in the national database in order to: 1) be consistent with all European countries, 2) be eligible for European comparisons and statistics, and 3) provide awareness on road fatalities and serious injuries data. Thus, the first suggestion to improve the Italian database is to collect and provide the crash severity on four levels of severity (Property Damage Only - PDO, slight injury, serious injury, and fatal). Moreover, the crash severity should be assessed through the use of the MAIS trauma scale, considering a value of MAIS equal to or greater than 3 to define as “serious injured” a person involved in the crash. This score should be evaluated by hospital reports. It is noteworthy to observe that the Italian highway police already collects PDO crashes but the Istat database does not report these crashes. As a consequence, money spent on PDO data collection is not used for crash prevention programs. Thus, inserting all crashes reported by the police in the database is highly recommended.

The information related to the alignment is collected by the police. However, the Istat form has a unique field both for crashes that occurred at intersections and not at intersections. This further implies that data related to horizontal alignment, vertical alignment, and intersection type are all collected under the name of the same variable even if actually they refer to different aspects of the roadway. Thus, a systematic collection of these data according to a specific format is strongly required.

Based on the comparison of the three national databases, a proposal for improvements of police crash data collection and for the Istat database format was formulated.

The comparison among the three databases pointed out that some new variables may be included in the Italian crash report form: 1) lighting (daytime, night-time, dawn/dusk), 2) day of the week (weekend or weekday), 3) speed limit, 4) crash location (segment or intersection), 5) alignment (curve or tangent), 6) intersection type (unsignalised, signalised, roundabout), 7) vehicle 1st point of impact, 8) vehicle engine (CC), 9) vehicle propulsion code, 10) vehicle age, and 11) safety equipment and devices (seat-belt, helmet).

Furthermore, the other national databases investigated in this work collect crash data by dividing general crash info, vehicle, and people involved-related information in different forms. This form allows flexibility in data recording. The Italian crash report form, instead, allows collecting at most three vehicles involved in the crash with a consistent loss of knowledge. Thus, the recommendation is to create three different forms to fulfil as for the British scheme:

1) A crash record form setting out the attendant circumstances associated with each crash, including information related to the roadway, the environment and the crash dynamics.

2) A vehicle record form for each of the vehicles involved in the crash, containing the vehicle's info and the driver-related data.

3) A casualty record for each casualty, containing information about the age and gender of the casualty and the severity of their injuries (fatal, serious, slight).

7.7 Conclusion, limitations, and outlook

In conclusion, the comparison among the performance metrics exhibited by all the models demonstrated that each group of methods has its pros and cons. However, the econometric models confirmed their advantages in offering easy to interpret outputs and understandable relations between dependent and independent variables whereas machine learning tools exhibited higher classification accuracy and ability to highlight hidden relations among data (a greater number of significant predictors identified by machine learning tools has been observed for each case studio reported in this research). The research results further suggest that the combined use of econometric methods and machine learning algorithms may overcome the limits of each group of methods with a satisfactory trade-off between prediction accuracy and interpretation of the results. Thus, in relation to the choice of the most eligible method to conduct crash severity analyses, the analyst may opt for the joint use of econometric methods and machine learning algorithms by exploiting:

- The interpretability of the econometric methods and
- The ability of the machine learning tools to create comprehensible scenarios (as those provided by association rules, classification tree, and random forest) to identify the co-occurrence of the patterns when a vehicle-pedestrian crash occurs or the straightforward ability in classification accuracy (as expressed by support vector machine and artificial neural network) in selecting the most

important predictors and providing powerful insights on the variables that should be better investigated in the analysis.

Then, downstream of the safety analysis, factors contributing to fatal and serious crashes can be identified. The availability of consistent data has an effect on safety management allowing a more detailed identification of these factors. Despite the fact that the crash occurrence is susceptible to many factors related to uncontrollable aspects, such as the weather, the risk of being involved in the most serious pedestrian road crashes can be reduced by road safety countermeasures. The results of this research may be of support to transportation planners, decision-makers, and local authorities to plan, design, operate, and manage a safer transportation system. Thus, detected the interdependences between contributory patterns and crash pedestrian involvement, a combination of engineering, behavioural, and management strategies, as well as appropriate safety countermeasures, can be provided and planned to effectively moderate pedestrian crash severity increasing the perceived safety of walking and contributing to the vision zero-deaths on road by 2050.

Some limitations of this dissertation research may be that the three databases are extremely different in sample sizes but they also differ in the information they stored in terms of data accurateness and number of explanatory variables. Thus, further research will be carried out on the same dataset with different sample sizes (i.e., Istat data at national and local levels) to further investigate how to vary the performances of the models with different sample sizes but in presence of the same information available. In so doing, the safety countermeasures can be appropriately identified accounting for the study context and be of major support for the local authorities.

As future directions, this research will further explore the methodological approach developed in this dissertation framework. With regards to the econometric models, as the implementation of the random parameter models provides evidence of the existence of the heterogeneity among data, an in-depth investigation of how this aspect affects crash severity is thus encouraged. Regarding the machine learning tools, as the use of these models is generally spreading, there is a growing need to find out more comprehensive results of SVM, ANN, and RF applications in order to be of greater support in crash severity analyses.

REFERENCES

Abay, K.A., 2013. Examining pedestrian-injury severity using alternative disaggregate models. Res. Transp. Econ. 43, 123-136, http://dx.doi.org/10.1016/j.retrec.2012.12.002.

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. Proc. ACM SIGMOD Int. Conf. Manag. Data, 207–216, https://doi.org/10.1145/170035.170072.

Agresti, A., 2002. Categorical Data Analysis 3rd edition. ISBN: 978-0-470-46363-5.

Ahmadi, A., Jahangiri, A., Berardi, V., and Machiani, S. G., 2018. Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods. J. Transp. Saf. Secur. 12(4), 522-546, https://doi.org/10.1080/19439962.2018.1505793.

Allison, P.D., 2008. Convergence Failures in Logistic Regression, Paper 360-2008. SAS Global Forum 2008, available at: http://people.vcu.edu/~dbandyop/BIOS625/Convergence_Logistic.pdf.

Anastasopoulos, P. C., and Mannering, F. L., 2011. An empirical assessment of fixed and random parameter logit models using crash-and non-crash-specific injury data. Acc. Anal. Prev. 43(3), 1140-1147, https://doi.org/10.1016/j.aap.2010.12.024.

Andreß, H. J., Golsch, K., Schmidt, A. W., 2013. Applied panel data analysis for economic and social surveys. Springer Science & Business Media, https://doi.org/10.1007/978-3-642-32914-2.

Association for the Advancement of Automotive Medicine (AAAM), 2005. Abbreviated Injury Scale (AIS), available at: https://www.aaam.org/abbreviated-injury-scale-ais/.

Austroads, 2020. Updating Austroads Pedestrian Planning and Design Guidance in line with International Best Practice, Austroads Project No. NTM6118. ISBN 978-1-922382-33-7.

Aziz, H.M.A., Ukkusuri, S.V., and Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Acc. Anal. Prev. 50, 1298-1309, http://dx.doi.org/10.1016/j.aap.2012.09.034.

Babić, D., Babić, D., Fiolić, M., Ferko, M., 2021. Factors affecting pedestrian conspicuity at night: Analysis based on driver eye tracking. Saf. Sci. 139, 105257, https://doi.org/10.1016/j.ssci.2021.105257.

Baker, S.P., O'Neill, B., Haddon, W., and Long, W.B., 1974. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J. Trauma 14(3), 187–196, PMID: 4814394.

Besharati, M. M., and Kashani, A.T., 2017. Which set of factors contribute to increase the likelihood of pedestrian fatality in road crashes?. Int. J. Inj. Control Saf. Promot. 25(3), 247-256, https://doi.org/10.1080/17457300.2017.1363781.

Bina, B., Schulte, O., Crawford, B., Qian, Z., Xiong, Y., 2013. Simple decision forests for multi-relational classification. Decis. Support Syst. 54(3), 1269-1279, https://doi.org/10.1016/j.dss.2012.11.017.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA.

Casado-Sanz, N., Guirao, B., Galera, A. L., and Attard, M., 2019. Investigating the Risk Factors Associated with the Severity of the Pedestrians Injured on Spanish Crosstown Roads. Sustainability 11(19), 5194, https://doi.org/10.3390/su11195194.

Cerwick, D.M., Gkritza, K., Shaheed, M.S., and Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Anal. Methods Accid. Res. 3(4), 11-27, http://dx.doi.org/10.1016/j.amar.2014.09.002.

Chawla, N. V., Bowyer, Hall, L. O., and Kegelmeyer W. P., 2002. Smote: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357, http://dx.doi.org/10.1613/jair.953.

Chawla, N. V., Japkowicz, N., and Kołcz, A., 2004. ACM SIGKDD Special Issue on Learning from Imbalanced Data Sets 6(1), 1-6, https://doi.org/10.1145/1007730.1007733.

Chen, Z. and Fan, W.D., 2019. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. Int. J. Transp. Sci. Technol. 8(1), 43-52, https://doi.org/10.1016/j.ijtst.2018.10.001.

Clifton, K.J., Burnier, C., and Akar, G., 2009. Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built environment?. Transp. Res. Part D 14(6), 425-436, https://doi.org/10.1016/j.trd.2009.01.001.

CrowdFlower, 2016. Data Science Report. http://www2.cs.uh.edu/~ceick/UDM/CFDS16.pdf.

Damju, J. S., Wening, B., Das, T., and Lee, D., 2020. Learning Spark: lightning-fast big data analytics, 2nd edn. O'Reilly Media, available at: https://pages.databricks.com/rs/094-YMS-629/images/LearningSpark2.0.pdf.

Das, S., Dutta, A., Dixon, K., Sun, X., and Jalayer, M., 2019. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. Int. J. Urban Sci. 23(1), 30-48, https://doi.org/10.1080/12265934.2018.1431146

de Villiers, J., and Barnard, E., 1993. Backpropagation neural nets with one and two hidden layers. IEEE Trans. Neural Netw. 4(1), 136–141, http://dx.doi.org/10.1109/72.182704.

Decree 2008-754 of 30th July 2008, France, pertaining to modifications to the Code de la route, available at: https://www.legifrance.gouv.fr/loda/id/LEGIARTI000019275395/2008-08-02/.

Demsere-Derry, J., Ebel, B., Mock, C.N., and Donkor, P., 2010. Pedestrians' injury patterns in Ghana. Acc. Anal. Prev. 42(4), 1080-1088, https://doi.org/10.1016/j.aap.2009.12.016.

Department of Transportation (DOT), 2012. MMUCC guideline model minimum uniform crash criteria, 4th Edition. DOT HS 811 631, available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811631.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Acc. Anal. Prev. 40, 1033-1054, https://doi.org/10.1016/j.aap.2007.11.010.

ETSC, 2021.15th Annual Road Safety Performance Index (PIN) Report. Background tables, available at:https://etsc.eu/15th-annual-road-safety-performance-index-pin-report/.

European Commission, 2017. Council conclusions on road safety - endorsing the Valletta Declaration of March 2017, available at: https://data.consilium.europa.eu/doc/document/ST-9994-2017-INIT/en/pdf.

European Commission, 2018. Traffic Safety Basic Facts on Pedestrians, available at: https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs20xx_pedestrians.pdf.

European Commission, 2019a. EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero", available at: https://ec.europa.eu/transport/sites/transport/files/legislation/swd20190283-roadsafety-vision-zero.pdf.

European Commission, 2019b. Handbook on the External Costs of Transport. https://ec.europa.eu/transport/themes/sustainable/studies/sustainable_it.

European Commission, 2019c. Guidelines for Developing and implementing a Sustainable Urban Mobility Plan, 2nd Edition, available at: https://www.eltis.org/sites/default/files/sump_guidelines_2019_interactive_document_1.pdf.

European Commission, 2019d. Directive (EU) 2019/1936 of the European Parliament and of the Council of 23 October 2019 amending Directive 2008/96/EC on road infrastructure safety management, available at: http://data.europa.eu/eli/dir/2019/1936/oj.

European Commission, 2020. EU Road Safety Policy Framework 2021 – 2030, Next steps towards "Vision Zero", available at: https://op.europa.eu/en/publication-detail/-/publication/d7ee4b58-4bc5-11ea-8aa5-01aa75ed71a1.

European Commission, 2021. Traffic fatalities collision matrix in the EU, available at: https://www.nrso.ntua.gr/european-commission-traffic-fatalities-collision-matrix-in-the-eu-november-2021/.

Fekety, D.K., Edewaard, D.E., Stafford Sewall, A.A., Tyrrell, R.A., 2016. Electroluminescent Materials Can Further Enhance the Nighttime Conspicuity of Pedestrians Wearing Retroreflective Materials. Hum. Factors 58(7), 976–985, https://doi.org/10.1177/0018720816651535.

Fernandez, A., Garcìa, S., Galar, M., Prati, R. C., Krawczyk, B., and Herrera, F., 2018. Learning from imbalanced data sets. Springer, ISBN 978-3-319-98073-7, https://doi.org/10.1007/978-3-319-98074-4.

Fiorentini, N. and Losa, M., 2020. Handling Imbalanced Data in Road Crash Severity Prediction by Machine Learning Algorithms. Infrastructures 5(7), 61, https://doi.org/10.3390/infrastructures5070061.

Fitzpatrick, K., Turner, S.M., Brewer, M., Carlson, P.J., Ullman, B., Trout, N.D., Park, E.S., Whitacre, J., Lalani, N., Lord, D., 2006. TCRP Report 112/NCHRP Report 562: Improving Pedestrian Safety at Unsignalised Crossings. TRB of the National Academies, Washington, D.C.

Flannagan, C., Mann, N., and Rupp, J., 2013. Measuring serious injuries in traffic crashes. Transp. Res. Board 92nd Annual Meeting, Washington, DC, Paper No. 13-4780.

Ganganwar, V., 2012. An overview of classification algorithms for imbalanced. IJETAE 2(4), pp. 42-47, ISSN 2250-2459.

Ghasedi, M., Sarfjoo, M., and Bargegol, I., 2021. Prediction and Analysis of the Severity and Number of Suburban Accidents Using Logit Model, Factor Analysis and Machine Learning: A case study in a developing country. SN Appl. Sci. 3(13), https://doi.org/10.1007/s42452-020-04081-3.

Guo X., Yin, Y., Dong, C., Yang, G., and G. Zhou., 2008. On the Class Imbalance Problem. Fourth International Conference on Natural Computation, https://doi.org/10.1109/ICNC.2008.871.

Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., Wang, W., 2021. Older Pedestrian Traffic Crashes Severity Analysis Based on an Emerging Machine Learning XGBoost. Sustainability 13, 926. https://doi.org/10.3390/su13020926.

Haleem, K., Alluri, P., and Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Acc. Anal. Prev. 81, 14-23, https://doi.org/10.1016/j.aap.2015.04.025.

Hanson, C. S., Noland, R. B., and Brown, C., 2013. The severity of pedestrian crashes: an analysis using Google Street View imagery. J. Transp. Geogr. 33, 42–53, http://dx.doi.org/10.1016/j.jtrangeo.2013.09.002.

He, H. and Garcia, E. A., 2009. Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 21(9), pp. 1263-1284, https://doi.org/10.1109/TKDE.2008.239.

Helvik, R., Vaa, T., Hoye, A., and Sorensen, M., 2009. The Handbook of Road Safety Measures: Second Edition. ISBN: 978-1-84855-250-0.

Imprialou, M. & Quddus, M., 2019. Crash data quality for road safety research: Current state and future directions. Acc. Anal. Prev. 130, 84-90, https://doi.org/10.1016/j.aap.2017.02.022.

IRTAD, 2020. ROAD SAFETY REPORT 2020 | SWEDEN. OECD available at: https://www.itf-oecd.org/sites/default/files/sweden-road-safety.pdf.

Islam, S., and Jones, S., 2014. Pedestrian at-fault crashes on rural and urban roadways in Alabama. Acc. Anal. Prev. 72, 267-276, https://doi.org/10.1016/j.aap.2014.07.003.

Islam, Z., Andel-Aty, M., Cai, Q., and Yuan, J., 2021. Crash data augmentation using variational autoencoder. Acc. Anal. Prev. 105950, https://doi.org/10.1016/j.aap.2020.105950.

Ivan, J.N. and Konduri, K.C., 2018. Crash Severity Methods. Lord, D. and Washington, S. (Ed.) Safe Mobility: Challenges, Methodology and Solutions (Transport and Sustainability, Vol. 11), Emerald Publishing Limited, Bingley, 325-350. https://doi.org/10.1108/S2044-994120180000011017.

Jobson, J.D., 1992a. Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design. Springer Texts in Statistics, https://doi.org/10.1007/978-1-4612-0955-3.

Jobson, J.D., 1992b. Applied multivariate data analysis: volume II: Categorical and Multivariate Methods. Springer Texts in Statistics, https://doi.org/10.1007/978-1-4612-0921-8.

Kashani A.T. and Mohaymany A.S., 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Saf. Sci., 49(10), 1314-1320. https://doi.org/10.1016/j.ssci.2011.04.019.

Kashani, A., Rabieyan, R., and Besharati, M., 2014. A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers. Journal of Safety Research, 51, 93-98, http://dx.doi.org/10.1016/j.jsr.2014.09.004.

Kim, J.K., Ulfarsson, G.F., Sarkar, V.N., Mannering, F.L., 2010. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Acc. Anal. Prev. 42, 1751-1758, https://doi.org/ 10.1016/j.aap.2010.04.016.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Political Anal, 9(2), 137–163, https://doi.org/10.1093/oxfordjournals.pan.a004868.

Komol, M.R., Hasan, M., Elhenawy, M., Yasmin, S., Masoud, M., and Rakotonirainy, A., 2021. Crash severity analysis of vulnerable road users using machine learning. PLoSONE 16(8):e0255828, https://doi.org/10.1371/journal.pone.0255828.

Lamm, R., Beck, A., Mailaender, T., Cafiso, S., and La Cava, G., 2007. How to Make Two-Lane Rural Roads Safer. Scientific Background and Guide for Practical Application. WIT Press / Computational Mechanics, ISBN-101845641566.

Lantieri, C., Costa, M., Vignali, V., Acerra, E.M., Marchetti, P., Simone, A., 2021. Flashing in-curb LEDs and beacons at unsignalised crosswalks and driver's visual attention to pedestrians during nighttime. Ergonomics 64(3), 330–341, https://doi.org/10.1080/00140139.2020.1834149.

Lee, C., and Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. Acc. Anal. Prev. 37, 775-786, https://doi.org/10.1016/j.aap.2005.03.019.

Li, D., Ranjitkar, P., Zhao, Y., Yi, H., and Rashidi, S., 2017. Analyzing pedestrian crash injury severity under different weather conditions. Traffic Inj. Prev. 18(4), 427-430, https://doi.org/10.1080/15389588.2016.1207762.

Loftus, J.R. and Taylor, J.E., 2014. A significance test for forward stepwise model selection. Cornell University, arXiv:1405.3920.

Long, J. S., 1997. Regression models for categorical and limited dependent variables. ISBN: 0803973748.

López, G., Abellán, J., Montella, A, de Oña, J., 2014. Patterns of Single-Vehicle Crashes on Two-Lane Rural Highways in Granada Province, Spain: In-Depth Analysis through Decision Rules. Transp. Res. Rec., 2432, 133-141, https://doi.org/10.3141/2432-16.

Maglogiannis, I.G., Karpouzis, K., Wallace, M., Soldatos, J., 2007. Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies. IOS Press, ISSN 0922-6389.

Malyshkina N.V. and Mannering, F.L., 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle crashes. Acc. Anal. Prev. 42, 131-139, https://doi.org/10.1016/j.aap.2009.07.013.

Mannering, F.L., Bhat, C. R., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Anal. Methods Accid. Res. 25, 100113, https://doi.org/10.1016/j.amar.2020.100113.

Mannering, F.L., Shankar, V., Bhat, C. R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16, http://dx.doi.org/10.1016/j.amar.2016.04.001.

Martin, J.L., and Wu, D., 2017. Pedestrian fatality and impact speed squared: Cloglog modeling from French national data. Traffic Injury Prevention 19(1), 94-101, https://doi.org/10.1080/15389588.2017.1332408.

McFadden, D., 1981. Structural Analysis of Discrete Data with Econometric Applications. ISBN: 9780262131599.

Menardi, G., and Torelli, N., 2012. Training and assessing classification rules with imbalanced data. Data Mining and Knowledge Discover, 28, 92–122, https://doi.org/10.1007/s10618-012-0295-5.

Meocci, M., Branzi, V., Martini, G., Arrighi, R., and Petrizzo, I., 2021. A Predictive Pedestrian Crash Model Based on Artificial Intelligence Techniques. Appl. Sci. 11, 11364, https://doi.org/10.3390/app112311364.

Milton, j., 2006. Highway accident severities and the mixed logit model: An exploratory empirical analysis. Acc. Anal. Prev. 40(1), 260-266, https://doi.org/10.1016/j.aap.2007.06.006.

Mokhtarimousavi, S., Anderson, J.C., Azizinamini, A., and Hadi, M., 2020. Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. IJTST 9, 100-115, https://doi.org/10.1016/j.ijtst.2020.01.001.

Montella, A., Aria, M., D'Ambrosio, A., and Mauriello, F., 2011. Data-Mining Techniques for Exploratory Analysis of Pedestrian Crashes. Trans. Res. Rec. 2237, 107-116, https://doi.org/10.3141/2237-12.

Montella, A., Aria, M., D'Ambrosio, A., and Mauriello, F., 2012a. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid. Anal. Prev. 134, 58-72, https://10.1016/j.aap.2011.04.025.

Montella, A., Mauriello, F., Tarko, A., Romero, M.A., 2012b. Critical Review of the International Crash Databases and Proposals for Improvement of the Italian National Database. Procedia - Social and Behavioral Sciences 53, 49-61, https://doi.org/10.1016/j.sbspro.2012.09.859.

Montella, A., Chiaradonna, S., Criscuolo, G., and De Martino, S., 2019. Development and evaluation of a web-based software for crash data collection, processing and analysis. Accid. Anal. Prev, http://dx.doi.org/10.1016/j.aap.2017.01.013.

Montella, A., de Oña, R., Mauriello, F., Rella Riccardi, M., Silvestro, G., 2020. A data mining approach to investigate patterns of powered two-wheeler crashes in Spain. Accid. Anal. Prev. 134, 105251, https://doi.org/10.1016/j.aap.2019.07.027.

Montella, A., Mauriello, F., Pernetti, M., Rella Riccardi, M., 2021. Rule discovery to identify patterns contributing to overrepresentation and severity of run-off-the-road crashes. Accid. Anal. Prev. 155, 106119, https://doi.org/10.1016/j.aap.2021.106119.

Moral-Garcia, S., Castellano, J.G., Mantas, J.G., Montella, A., Abellan, J., 2019. Decision tree ensemble method for analyzing traffic accidents of novice drivers in urban areas. Entropy 21(4), 360, https://doi.org/10.3390/e21040360

Moudon, A.V., Lin, L., Jiao, J., Hurvitz, P., and Reeves, P., 2011. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. Acc. Anal. Prev. 43, 11-24, https://doi:10.1016/j.aap.2009.12.008.

National Highway Traffic Safety Administration - NHTSA, 2017. MMUCC Guideline Model Minimum Uniform Crash Criteria, 5th Edition.

National Technical University of Athens, 2011. Strategic Plan for the improvement of road safety in Greece, 2011-2020, available at: https://ec.europa.eu/transport/road_safety/sites/default/files/pdf/strategic_plan_for_road_safety_greece_2011_2020.pdf.

Ndour, C., Diop, A., Dossou-Gbété, S., 2012. Classification approach based on association rules mining for unbalanced data, available at: https://arxiv.org/pdf/1202.5514.pdf.

Noh, Y., Kim, M., Yoon, Y., 2019. Elderly pedestrian safety in a rapidly aging society—Commonality and diversity between the younger-old and older-old. Traffic Injury Prevention 19(8), 874-879, https://doi.org/10.1080/15389588.2018.1509209.

OECD, 2012. Pedestrian safety, urban space and health. Organisation for Economic Co-operation and Development OECD, Paris, available at: https://read.oecd-ilibrary.org/transport/pedestrian-safety-urban-space-and-health_9789282103654-en#page15.

Olszewski, P., Szagala, P., Wolanski, M., and Zielinska, A., 2015. Pedestrian fatality risk in accidents at unsignalized zebra crosswalksin Poland. Acc. Anal. Prev. 84, 83-91, http://dx.doi.org/10.1016/j.aap.2015.08.008.

Ospina-Mateus, H., Quintana Jiménez, L.A., López-Valdés, F.J., Morales-Londoño, N., and Salas-Navarro, K., 2019. Using Data-Mining Techniques for the Prediction of the Severity of Road Crashes in Cartagena, Colombia. WEA 2019: Applied Computer Sciences in Engineering, 309-320, https://doi.org/10.1007/978-3-030-31019-6_27.

Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Acc. Anal. Prev. 42, 1839-1854, https://doi.org/10.1016/j.aap.2010.05.005.

Pour-Rouholamin, M. and Zhou, H., 2016. Investigating the risk factors associated with pedestrian injury severity in Illinois. J. Saf. Res. 57, 9-17, https://doi.org/10.1016/j.jsr.2016.03.004.

Rella Riccardi, M., Mauriello, F., Sarkar, S, Montella, A., 2022. Analysing factors contributing to severe and fatal pedestrian crashes in Great Britain by econometric models and machine learning algorithms. Sustainability 14(6), 3188, https://doi.org/10.3390/su14063188.

Rothman, L., Howard, A. W., Camden, A., and Macarthur, C., 2015. Pedestrian crossing location influences injury severity in urban areas. Injury Prevention 18, 365–370, http://dx.doi.org/10.1136/injuryprev-2011-040246.

Sáez, J.A., Luengo, J., Stefanowski, J., and Herrera, F., 2015. SMOTE--IPF: addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. Inf. Sci., 291, 184-203, https://doi.org/10.1016/j.ins.2014.08.051.

Savolainen, P., Mannering, F., Lord, D., and Quddus, M., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Acc. Anal. Prev. 43, 1666-1676, https://doi.org/10.1016/j.aap.2011.03.025.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., and Milton, J. (2017). Occupant injury severities in hybrid-vehicle involved crashes: a random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research, 15, 41-55, https://doi.org/10.1016/j.amar.2017.05.003.

Singh, K., 2020. How to Improve Class Imbalance using Class Weights in Machine Learning, available at: https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/.

Sivasankaran, S., K., Natarajan, P., and Balasubramanian, V., 2020. Identifying Patterns of Pedestrian Crashes in Urban Metropolitan Roads in India using Association Rule Mining. Transp. Res. Procedia 48, 3496-3507, https://doi.org/10.1016/j.trpro.2020.08.102.

Srinivasan, K. K., 2002. Injury Severity Analysis with Variable and Correlated Thresholds: Ordered Mixed Logit Formulation. Trans. Res. Rec. 1784, 132-142, https://doi.org/10.3141%2F1784-17.

Strickland, J., 2017. Data science applications using R. ISBN: 978-0-359-81042-0.

Theofilatos, A., and Yannis, G., 2015. A review of powered-two-wheeler behaviour and safety. International Journal of Injury Control and Safety Promotion 22(4), 284 – 307, https://doi.org/10.1080/17457300.2014.908224.

Tay, R., Choi, J., Kattan, L., and Khan, A., 2011. A Multinomial Logit Model of Pedestrian–Vehicle Crash Severity. Int. J. Sustain. Transp. 5, 233-249, https://doi.org/10.1080/15568318.2010.497547.

Train, K., 2009. Discrete choice Methods with simulation, 2nd edition. ISBN 978-0-521-76655-5.

Tulu, G.S., Washington, S., Haque, M., and King, M., 2017. Injury severity of pedestrians involved in road traffic crashes in Addis Ababa, Ethiopia. J. Transp. Saf. 9(1), 47-66, https://doi.org/10.1080/19439962.2016.1199622.

United Nations, 1968. Convention on Road Traffic. Vienna 8 November 1968, Vol.1, Chap. XI, available at: https://treaties.un.org/doc/Treaties/1977/05/19770524%2000-13%20AM/Ch_XI_B_19.pdf.

United Nations, 2016. Resolution 70/260 Improving global road safety, available at: https://undocs.org/pdf?symbol=en/A/RES/70/260.

United Nations, 2021. Global Plan for the Decade of Action for Road Safety 2021-2030, available at: https://www.who.int/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2021-2030.

Washington, S. P., Karlaftis, M. G., Mannering, F. L., 2011. Statistical And Econometric Methods for Transportation Data Analysis, 2nd edition. ISBN 9781420082852.

Watson, A., Watson, B., and Vallmuur, K., 2015. Estimating under-reporting of road crash injuries to police usingmultiple linked data collections. Acc. Anal. Prev. 83, 10-25, http://dx.doi.org/10.1016/j.aap.2015.06.011.

Weijermars, W., Bos, N., Schoeters, A., Meunier, L.C., Nuyttens, N., Dupont, E., Machata, K., Bauer, R., Perez, K., Martin, J.L., Johansson, H., Filtness, A.,Brown, L., and Thomas, P., 2018. Serious Road Traffic Injuries in Europe, Lessons from the EU Research Project SafetyCube. Trans. Res: Rec., 2672(32), 1-9, https://doi.org/10.1177/0361198118758055.

Wittink, R., 2001. Promotion of mobility and safety of vulnerable road users. Final report of the European research project PROMISING, D-2001-3, available at: https://www.swov.nl/sites/default/files/publicaties/rapport/d-2001-03.pdf.

Wood, J.M., Tyrrell, R.A., Lacherez, P., Black, A.A., 2017. Night-time Pedestrian Conspicuity: Effects of Clothing on Drivers' Eye Movements. Ophthalmic Physiol Opt. 37(2), 184–190, https://doi.org/10.1111/opo.12351.

Word Health Organization, 2018. Global Status Report on Road Safety 2018. Licence: CC BY NC- SA 3.0 IGO, available at: https://www.who.int/publications/i/item/9789241565684.

Yamamoto, T., Hashiji, J., Shankar, N., 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Acc. Anal. Prev. 40(4), 1320-1329, https://doi.org/10.1016/j.aap.2007.10.016.

Yasmin, S., Eluru, S., Ukkusuri, S., 2014. Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City. J. Transp. Saf. Secur. 6(4), 275-300, https://doi.org/10.1080/19439962.2013.839590.

Ye, F. and Lord, D., 2011. Investigating the Effects of Underreporting of Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit and Mixed Logit Models. Transp. Res. Rec. 2241, 51-58, http://dx.doi.org/10.3141/2241-06.

Zeng, Q., Huang, H., Pei, X., Wong, S.C., and Gao, M., 2016. Rule extraction from an optimized neural network for traffic crash frequency modelling. Acc. Anal. Prev. 97, 87–95, http://dx.doi.org/10.1016/j.aap.2016.08.017.

Zhai, X., Huang, H., Sze, N.N., Song, Z., and Hon, K.K., 2019. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. Acc. Anal. Prev. 122, 3118-324, https://doi.org/10.1016/j.aap.2018.10.017.

Zhang, G., Yau, K.K.W., and Zhang, X., 2014. Analyzing fault and severity in pedestrian–motor vehicle accidents in China. Acc. Anal. Prev. 73, 141-150, http://dx.doi.org/10.1016/j.aap.2014.08.018.

Zhang, J., Li, Z., Pu, Z., and Xu, C., 2018. Comparing Prediction Performance for Crash Injury Severity among Various Machine Learning and Statistical Methods. IEEE Access 6, 60079-60087, 10.1109/ACCESS.2018.2874979.

APPENDIX 1 ~ GREAT BRITAIN

Classification tree

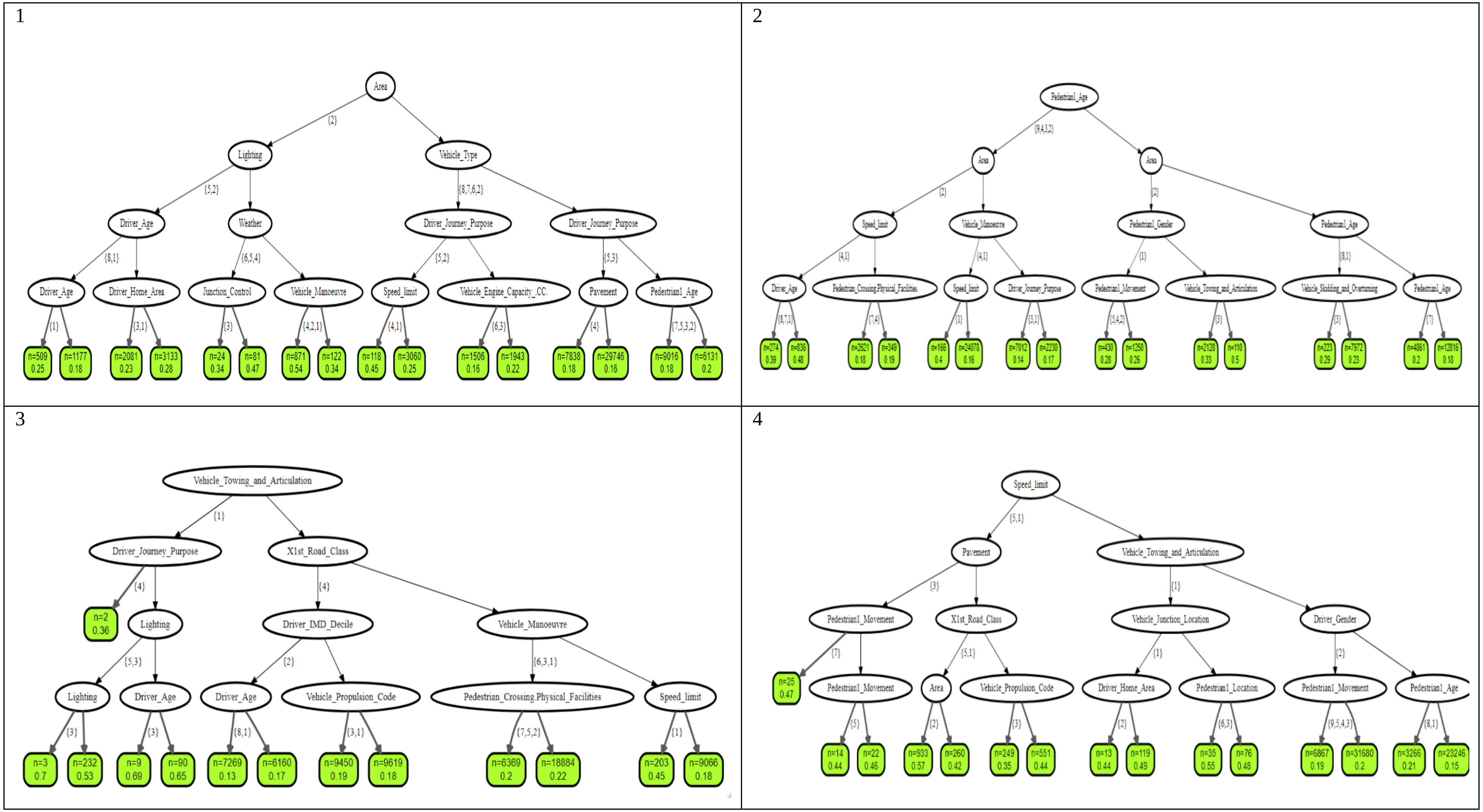
Table 143 – Tree in table format, Great Britain.

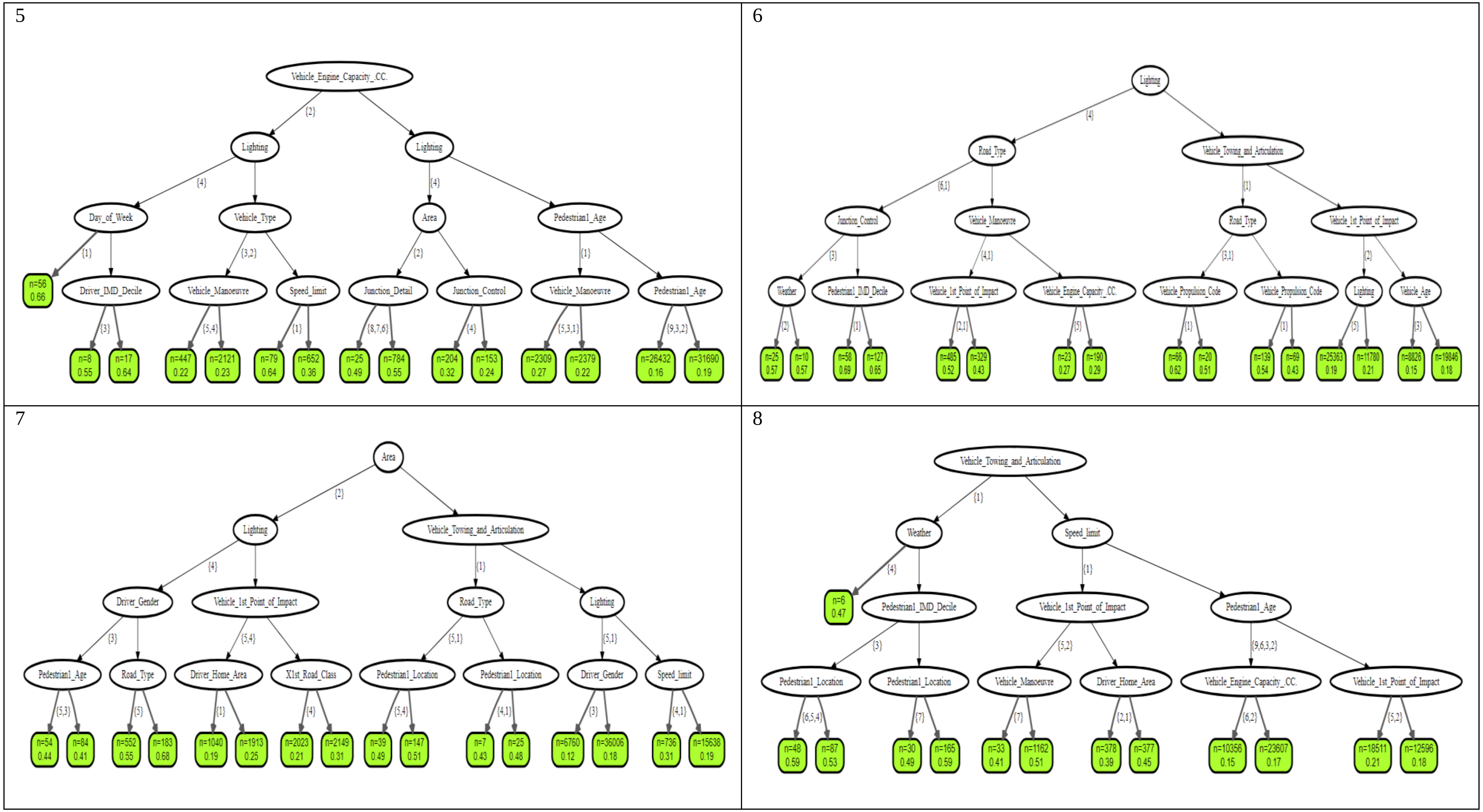
Node	Fatal		Serious		Slight		Total		Predicted Category	Parent Node
	N	%	N	%	N	%	N	%		
0	1,366	2.0	16,359	24.3	49,631	73.7	67,356	100.0	Slight	
1	471	10.6	1,510	34.1	2,449	55.3	4,430	16.2	Fatal	0
2	895	1.4	14,849	23.6	47,182	75.0	62,926	83.8	Slight	0
3	129	5.4	829	34.7	1,429	59.9	2,387	5.8	Fatal	1
4	342	16.7	681	33.3	1,020	49.9	2,043	10.4	Fatal	1
5	447	4.9	3,198	35.4	5,393	59.7	9,038	21.0	Fatal	2
6	448	0.8	11,651	21.6	41,789	77.5	53,888	62.7	Slight	2
7	124	6.5	683	35.8	1,100	57.7	1907	5.2	Fatal	3
8	5	1.0	146	30.4	329	68.5	480	.6	Serious	3
9	94	8.3	357	31.7	675	59.9	1,126	3.5	Fatal	4
10	248	27.0	324	35.3	345	37.6	917	6.9	Fatal	4
11	326	7.3	1,642	36.9	2,483	55.8	4,451	13.0	Fatal	5
12	121	2.6	1,556	33.9	2,910	63.4	4,587	8.1	Serious	5
13	199	0.5	7,563	20.1	29,942	79.4	37,704	40.4	Slight	6
14	249	1.5	4,088	25.3	11,847	73.2	16,184	22.4	Serious	6
15	18	3.0	197	32.5	391	64.5	606	1.1	Fatal	7
16	106	8.1	486	37.4	709	54.5	1,301	4.1	Fatal	7
17	38	26.6	42	29.4	63	44.1	143	1.1	Fatal	9
18	56	5.7	315	32.0	612	62.3	983	2.4	Fatal	9
19	219	35.4	232	37.5	168	27.1	619	5.9	Fatal	10
20	29	9.7	92	30.9	177	59.4	298	1.0	Fatal	10
21	92	4.2	739	33.9	1,349	61.9	2,180	4.7	Fatal	11
22	234	10.3	903	39.8	1,134	49.9	2,271	8.3	Fatal	11
23	88	2.1	1,446	34.4	2,668	63.5	4,202	6.9	Serious	12
24	33	8.6	110	28.6	242	62.9	385	1.2	Fatal	12
25	31	4.2	158	21.6	544	74.2	733	1.4	Fatal	13
26	168	0.5	7,405	20.0	29,398	79.5	36,971	38.9	Slight	13
27	212	2.1	2,990	29.0	7,097	68.9	10,299	16.0	Serious	14
28	37	0.6	1,098	18.7	4,750	80.7	5,885	6.3	Slight	14

Table 144 – Posterior Classification Ratio (PCR) for all nodes, Great Britain.

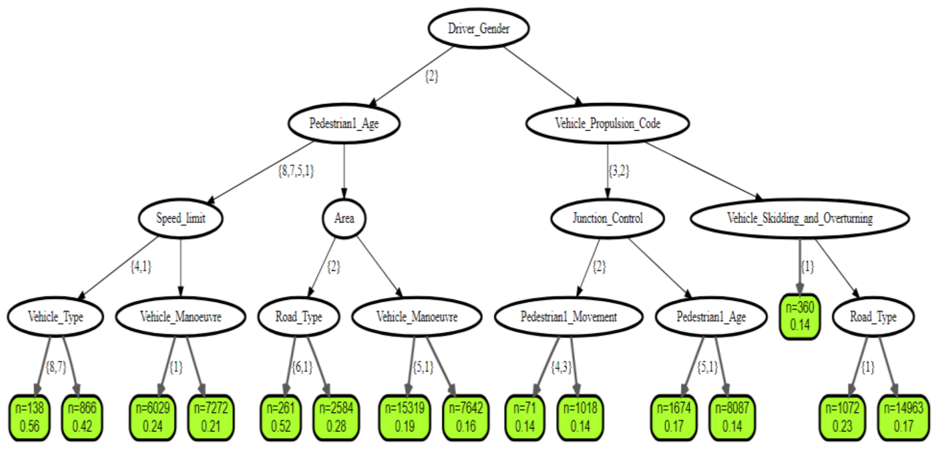
Node	PCR			
	Fatal	Serious	Slight	Actual Predicted Class
0	1.00	1.00	1.00	-
1	5.24	1.40	0.75	Fatal
2	0.70	0.97	1.02	Slight
3	2.66	1.43	0.81	Fatal
4	8.25	1.37	0.68	Fatal
5	2.44	1.46	0.81	Fatal
6	0.41	0.89	1.05	Slight
7	3.21	1.47	0.78	Fatal
8	0.51	1.25	0.93	Serious
9	4.12	1.31	0.81	Fatal
10	13.34	1.45	0.51	Fatal
11	3.61	1.52	0.76	Fatal
12	1.30	1.40	0.86	Serious
13	0.26	0.83	1.08	Slight
14	0.76	1.04	0.99	Serious
15	1.46	1.34	0.88	Fatal
16	4.02	1.54	0.74	Fatal
17	13.10	1.21	0.60	Fatal
18	2.81	1.32	0.84	Fatal
19	17.45	1.54	0.37	Fatal
20	4.80	1.27	0.81	Fatal
21	2.08	1.40	0.84	Fatal
22	5.08	1.64	0.68	Fatal
23	1.03	1.42	0.86	Serious
24	4.23	1.18	0.85	Fatal
25	2.09	0.89	1.01	Fatal
26	0.22	0.82	1.08	Slight
27	1.02	1.20	0.94	Serious
28	0.31	0.77	1.10	Slight

Random forest

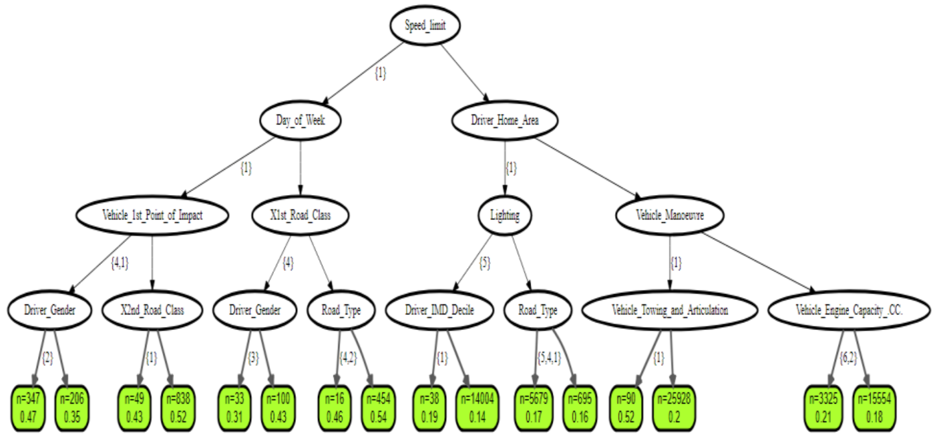



5
Vehicle_Engine_Capacity_CC.
{2}
Lighting
Lighting
{4}
{4}
Day_of_Week
Vehicle_Type
Area
Pedestrian1_Age
{1}
{3,2}
{2}
{1}
n=56 0.66
Driver_IMD_Decile
Vehicle_Manoeuvre
Speed_limit
Junction_Detail
Junction_Control
Vehicle_Manoeuvre
Pedestrian1_Age
{3}
{5,4}
{1}
{8,7,6}
{4}
{5,3,1}
{9,3,2}
n=8 0.55
n=17 0.64
n=447 0.22
n=2121 0.23
n=79 0.64
n=652 0.36
n=25 0.49
n=784 0.55
n=204 0.32
n=153 0.24
n=2309 0.27
n=2379 0.22
n=26432 0.16
n=31690 0.19
6
Lighting
{4}
Road_Type
Vehicle_Towing_and_Articulation
{6,1}
{1}
Junction_Control
Vehicle_Manoeuvre
Road_Type
Vehicle_1st_Point_of_Impact
{3}
{4,1}
{3,1}
{2}
Weather
Pedestrian1_IMD_Decile
Vehicle_1st_Point_of_Impact
Vehicle_Engine_Capacity_CC.
Vehicle_Propulsion_Code
Vehicle_Propulsion_Code
Lighting
Vehicle_Age
{2}
{1}
{2,1}
{5}
{1}
{1}
{5}
{3}
n=25 0.57
n=10 0.57
n=58 0.69
n=127 0.65
n=485 0.52
n=329 0.43
n=23 0.27
n=190 0.29
n=66 0.62
n=20 0.51
n=139 0.54
n=69 0.43
n=25363 0.19
n=11780 0.21
n=8826 0.15
n=19846 0.18
7
Area
{2}
Lighting
Vehicle_Towing_and_Articulation
{4}
{1}
Driver_Gender
Vehicle_1st_Point_of_Impact
Road_Type
Lighting
{3}
{5,4}
{5,1}
{5,1}
Pedestrian1_Age
Road_Type
Driver_Home_Area
X1st_Road_Class
Pedestrian1_Location
Pedestrian1_Location
Driver_Gender
Speed_limit
{5,3}
{5}
{1}
{4}
{5,4}
{4,1}
{3}
{4,1}
n=54 0.44
n=84 0.41
n=552 0.55
n=183 0.68
n=1040 0.19
n=1913 0.25
n=2023 0.21
n=2149 0.31
n=39 0.49
n=147 0.51
n=7 0.43
n=25 0.48
n=6760 0.12
n=36006 0.18
n=736 0.31
n=15638 0.19
8
Vehicle_Towing_and_Articulation
{1}
Weather
Speed_limit
{4}
{1}
n=6 0.47
Pedestrian1_IMD_Decile
Vehicle_1st_Point_of_Impact
Pedestrian1_Age
{3}
{5,2}
{9,6,3,2}
Pedestrian1_Location
Pedestrian1_Location
Vehicle_Manoeuvre
Driver_Home_Area
Vehicle_Engine_Capacity_CC.
Vehicle_1st_Point_of_Impact
{6,5,4}
{7}
{7}
{2,1}
{6,2}
{5,2}
n=48 0.59
n=87 0.53
n=30 0.49
n=165 0.59
n=33 0.41
n=1162 0.51
n=378 0.39
n=377 0.45
n=10356 0.15
n=23607 0.17
n=18511 0.21
n=12596 0.18

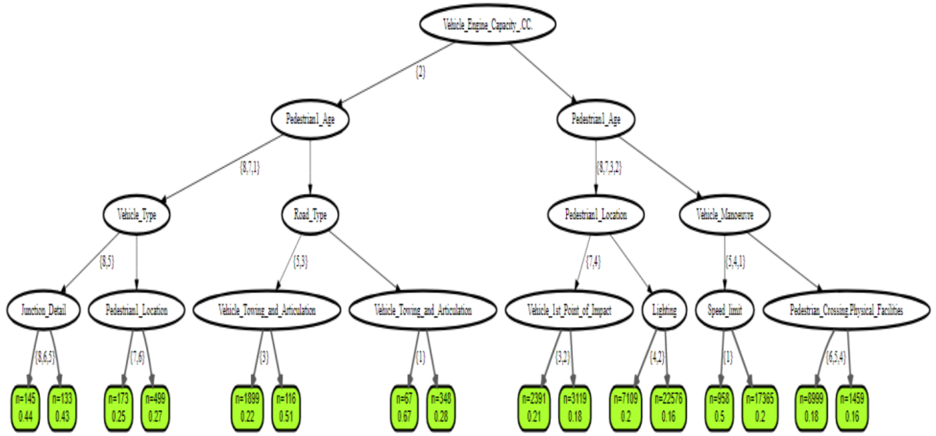
9


Driver_Gender
{2}
Pedestrian1_Age
Vehicle_Propulsion_Code
{8,7,5,1}
{3,2}
Speed_limit
Area
Junction_Control
Vehicle_Skidding_and_Overturning
{4,1}
{2}
{2}
{1}
Vehicle_Type
Vehicle_Manoeuvre
Road_Type
Vehicle_Manoeuvre
Pedestrian1_Movement
Pedestrian1_Age
n=360 0.14
Road_Type
{8,7}
{1}
{6,1}
{5,1}
{4,3}
{5,1}
{1}
n=138 0.56
n=866 0.42
n=6029 0.24
n=7272 0.21
n=261 0.52
n=2584 0.28
n=15319 0.19
n=7642 0.16
n=71 0.14
n=1018 0.14
n=1674 0.17
n=8087 0.14
n=1072 0.23
n=14963 0.17

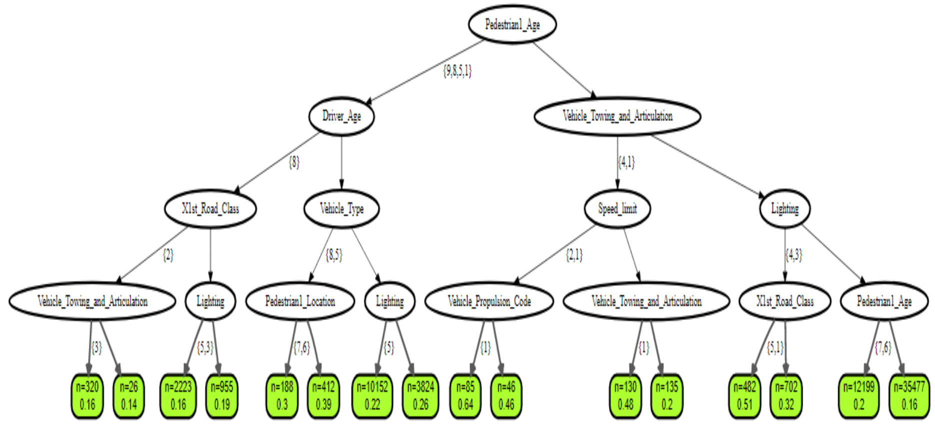
10


Speed_limit
{1}
Day_of_Week
Driver_Home_Area
{1}
{1}
Vehicle_1st_Point_of_Impact
X1st_Road_Class
Lighting
Vehicle_Manoeuvre
{4,1}
{4}
{5}
{1}
Driver_Gender
X2nd_Road_Class
Driver_Gender
Road_Type
Driver_IMD_Decile
Road_Type
Vehicle_Towing_and_Articulation
Vehicle_Engine_Capacity_CC.
{2}
{1}
{3}
{4,2}
{1}
{5,4,1}
{1}
{6,2}
n=347 0.47
n=206 0.35
n=49 0.43
n=838 0.52
n=33 0.31
n=100 0.43
n=16 0.46
n=454 0.54
n=38 0.19
n=14004 0.14
n=5679 0.17
n=695 0.16
n=90 0.52
n=25928 0.2
n=3325 0.21
n=15554 0.18

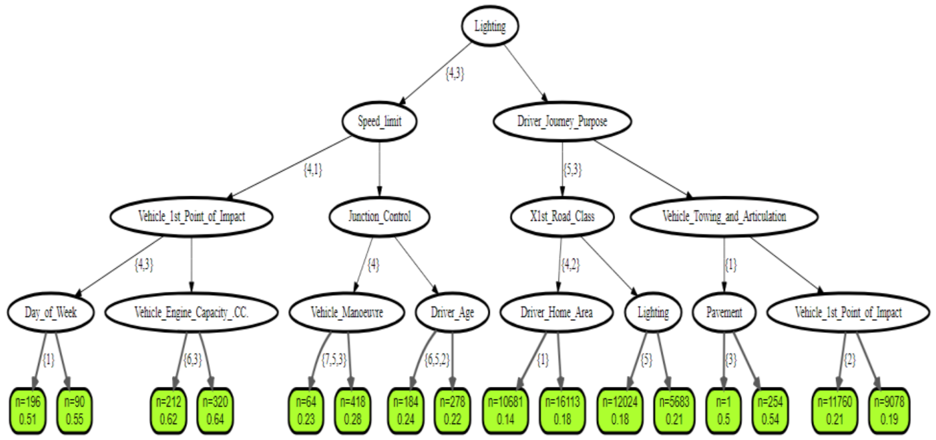
11


Vehicle_Engine_Capacity_CC.
{2}
Pedestrian1_Age
Pedestrian1_Age
{8,7,1}
{8,7,3,2}
Vehicle_Type
Road_Type
Pedestrian1_Location
Vehicle_Manoeuvre
{8,5}
{5,3}
{7,4}
{5,4,1}
Junction_Detail
Pedestrian1_Location
Vehicle_Towing_and_Articulation
Vehicle_Towing_and_Articulation
Vehicle_1st_Point_of_Impact
Lighting
Speed_limit
Pedestrian_Crossing.Physical_Facilities
{8,6,5}
{7,6}
{3}
{1}
{3,2}
{4,2}
{1}
{6,5,4}
n=145 0.44
n=133 0.43
n=173 0.25
n=499 0.27
n=1899 0.22
n=116 0.51
n=67 0.67
n=348 0.28
n=2391 0.21
n=3119 0.18
n=7109 0.2
n=22576 0.16
n=858 0.5
n=17365 0.2
n=8899 0.18
n=1459 0.16

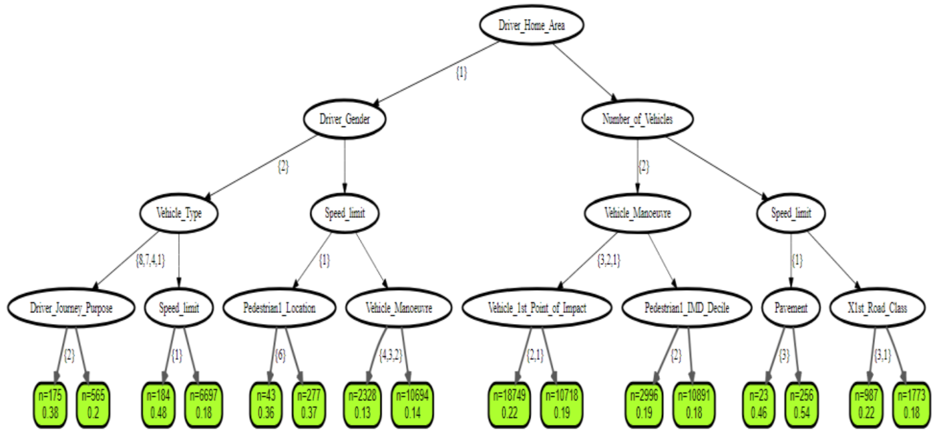
12


Pedestrian1_Age
{9,8,5,1}
Driver_Age
Vehicle_Towing_and_Articulation
{8}
{4,1}
X1st_Road_Class
Vehicle_Type
Speed_limit
Lighting
{2}
{8,5}
{2,1}
{4,3}
Vehicle_Towing_and_Articulation
Lighting
Pedestrian1_Location
Lighting
Vehicle_Propulsion_Code
Vehicle_Towing_and_Articulation
X1st_Road_Class
Pedestrian1_Age
{3}
{5,3}
{7,6}
{5}
{1}
{1}
{5,1}
{7,6}
n=320 0.16
n=26 0.14
n=2223 0.16
n=955 0.19
n=188 0.3
n=412 0.39
n=10152 0.22
n=3824 0.26
n=85 0.64
n=46 0.46
n=130 0.48
n=135 0.2
n=482 0.51
n=702 0.32
n=12199 0.2
n=35477 0.16

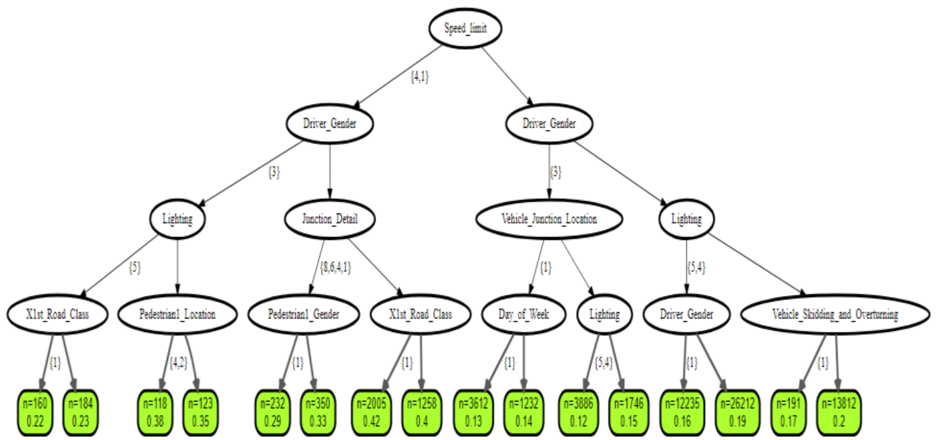
13


Lighting
{4,3}
Speed_limit
Driver_Journey_Purpose
{4,1}
{5,3}
Vehicle_1st_Point_of_Impact
Junction_Control
X1st_Road_Class
Vehicle_Towing_and_Articulation
{4,3}
{4}
{4,2}
{1}
Day_of_Week
Vehicle_Engine_Capacity_CC.
Vehicle_Manoeuvre
Driver_Age
Driver_Home_Area
Lighting
Pavement
Vehicle_1st_Point_of_Impact
{1}
{6,3}
{7,5,3}
{6,5,2}
{1}
{5}
{3}
{2}
n=196 0.51
n=90 0.55
n=212 0.62
n=320 0.64
n=64 0.23
n=418 0.28
n=184 0.24
n=278 0.22
n=10681 0.14
n=16113 0.18
n=12024 0.18
n=5683 0.21
n=1 0.5
n=254 0.54
n=11760 0.21
n=9078 0.19

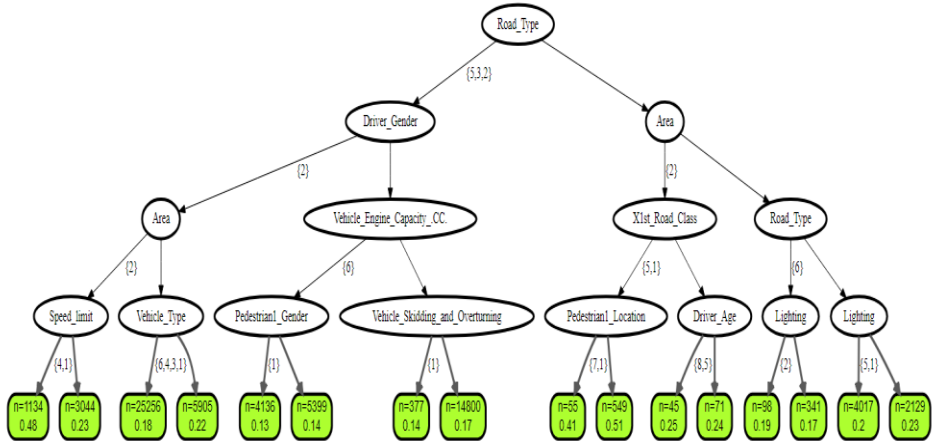
14


Driver_Home_Area
{1}
Driver_Gender
Number_of_Vehicles
{2}
{2}
Vehicle_Type
Speed_limit
Vehicle_Manoeuvre
Speed_limit
{8,7,4,1}
{1}
{3,2,1}
{1}
Driver_Journey_Purpose
Speed_limit
Pedestrian1_Location
Vehicle_Manoeuvre
Vehicle_1st_Point_of_Impact
Pedestrian1_IMD_Decile
Pavement
X1st_Road_Class
{2}
{1}
{6}
{4,3,2}
{2,1}
{2}
{3}
{3,1}
n=175 0.38
n=565 0.2
n=184 0.48
n=6697 0.18
n=43 0.36
n=277 0.37
n=2328 0.13
n=10694 0.14
n=18749 0.22
n=10718 0.19
n=2996 0.19
n=10891 0.18
n=23 0.46
n=256 0.54
n=987 0.22
n=1773 0.18

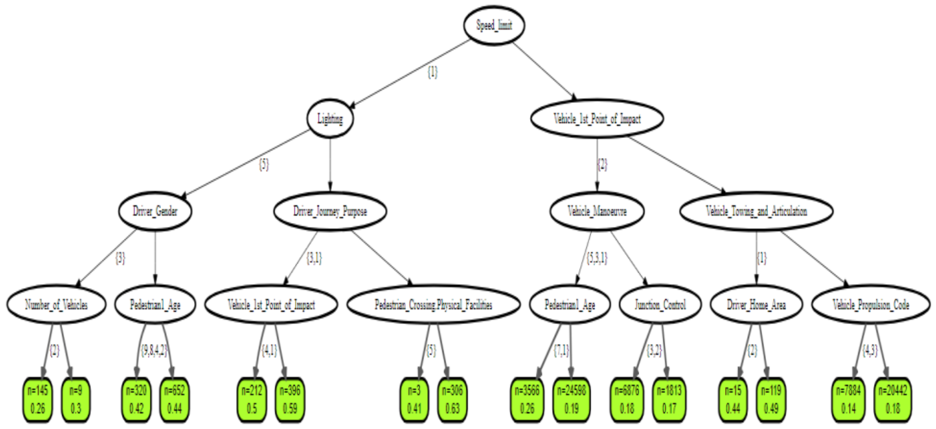
15


Speed_limit
{4,1}
Driver_Gender
Driver_Gender
{3}
{3}
Lighting
Junction_Detail
Vehicle_Junction_Location
Lighting
{5}
{8,6,4,1}
{1}
{5,4}
X1st_Road_Class
Pedestrian1_Location
Pedestrian1_Gender
X1st_Road_Class
Day_of_Week
Lighting
Driver_Gender
Vehicle_Skidding_and_Overturning
{1}
{4,2}
{1}
{1}
{1}
{5,4}
{1}
{1}
n=160 0.22
n=184 0.23
n=118 0.38
n=123 0.35
n=232 0.29
n=350 0.33
n=2005 0.42
n=1258 0.4
n=3612 0.13
n=1232 0.14
n=3886 0.12
n=1746 0.15
n=12235 0.16
n=26212 0.19
n=191 0.17
n=13812 0.2

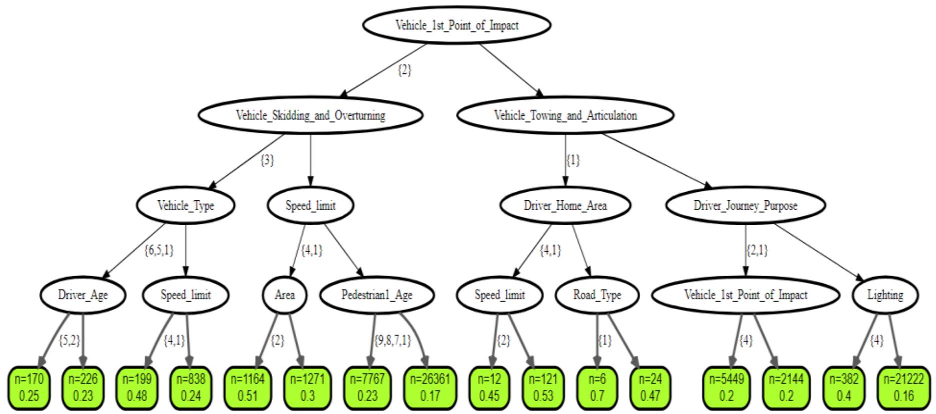
16


Road_Type
{5,3,2}
Driver_Gender
Area
{2}
{2}
Area
Vehicle_Engine_Capacity_CC.
X1st_Road_Class
Road_Type
{2}
{6}
{5,1}
{6}
Speed_limit
Vehicle_Type
Pedestrian1_Gender
Vehicle_Skidding_and_Overturning
Pedestrian1_Location
Driver_Age
Lighting
Lighting
{4,1}
{6,4,3,1}
{1}
{1}
{7,1}
{8,5}
{2}
{5,1}
n=1134 0.48
n=3044 0.23
n=25256 0.18
n=5906 0.22
n=4136 0.13
n=5399 0.14
n=377 0.14
n=14800 0.17
n=55 0.41
n=549 0.51
n=45 0.25
n=71 0.24
n=98 0.19
n=341 0.17
n=4017 0.2
n=2129 0.23

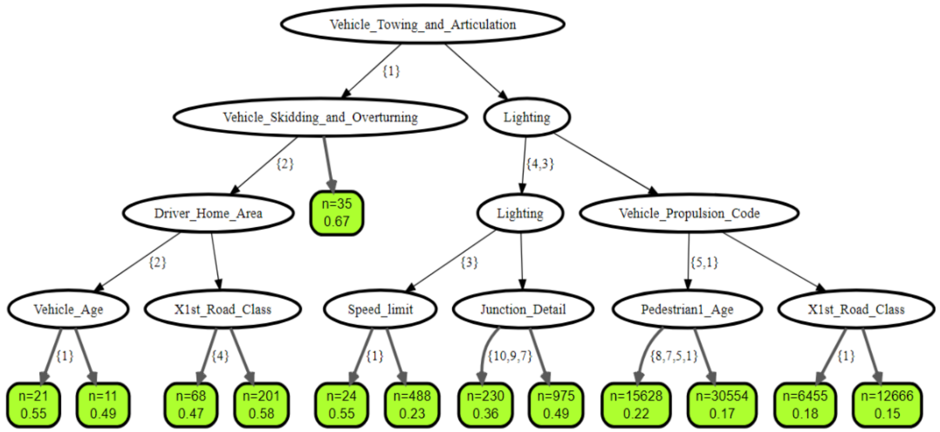
17


Speed_limit
{1}
Lighting
Vehicle_1st_Point_of_Impact
{5}
{2}
Driver_Gender
Driver_Journey_Purpose
Vehicle_Manoeuvre
Vehicle_Towing_and_Articulation
Number_of_Vehicles
Pedestrian1_Age
Vehicle_1st_Point_of_Impact
Pedestrian_Crossing.Physical_Facilities
Pedestrian1_Age
Junction_Control
Driver_Home_Area
Vehicle_Propulsion_Code

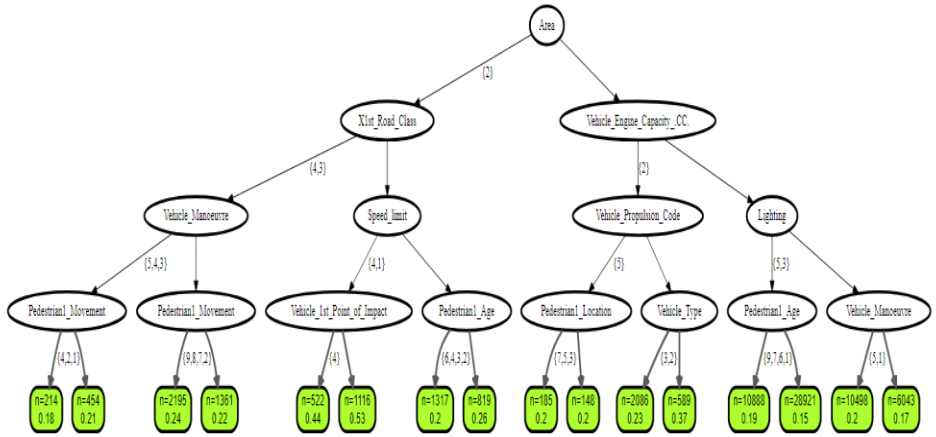
18


Vehicle_1st_Point_of_Impact
{2}
Vehicle_Skidding_and_Overturning
Vehicle_Towing_and_Articulation
{3}
{1}
Vehicle_Type
Speed_limit
Driver_Home_Area
Driver_Journey_Purpose
{6,5,1}
{4,1}
{4,1}
{2,1}
Driver_Age
Speed_limit
Area
Pedestrian1_Age
Speed_limit
Road_Type
Vehicle_1st_Point_of_Impact
Lighting
{5,2}
{4,1}
{2}
{9,8,7,1}
{2}
{1}
{4}
{4}
n=170
0.25
n=226
0.23
n=199
0.48
n=838
0.24
n=1164
0.51
n=1271
0.3
n=7767
0.23
n=26361
0.17
n=12
0.45
n=121
0.53
n=6
0.7
n=24
0.47
n=5449
0.2
n=2144
0.2
n=382
0.4
n=21222
0.16

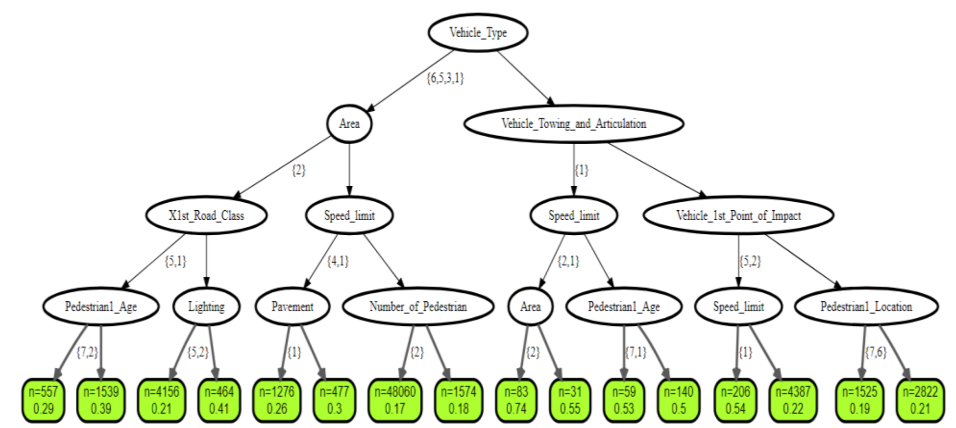
19


Vehicle_Towing_and_Articulation
{1}
Vehicle_Skidding_and_Overturning
Lighting
{2}
{4,3}
Driver_Home_Area
n=35
0.67
Lighting
Vehicle_Propulsion_Code
{2}
{3}
{5,1}
Vehicle_Age
X1st_Road_Class
Speed_limit
Junction_Detail
Pedestrian1_Age
X1st_Road_Class
{1}
{4}
{1}
{10,9,7}
{8,7,5,1}
{1}
n=21
0.55
n=11
0.49
n=68
0.47
n=201
0.58
n=24
0.55
n=488
0.23
n=230
0.36
n=975
0.49
n=15628
0.22
n=30554
0.17
n=6455
0.18
n=12666
0.15

20

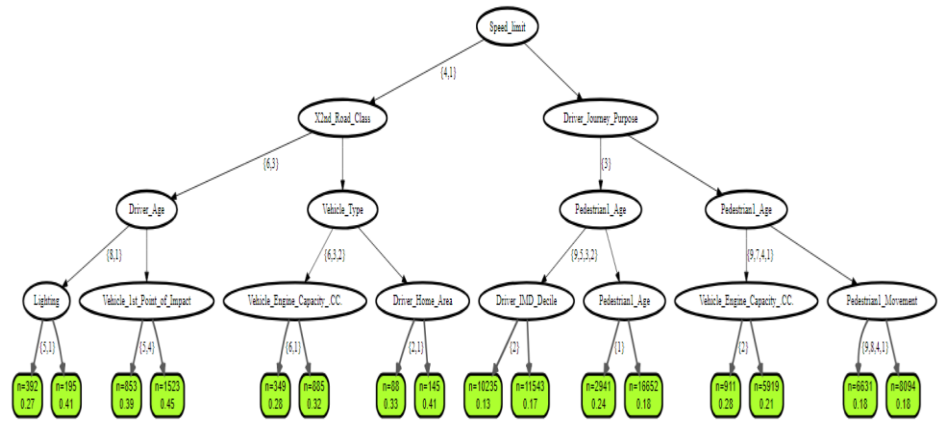

Area
{2}
X1st_Road_Class
Vehicle_Engine_Capacity_.CC.
{4,3}
{2}
Vehicle_Manoeuvre
Speed_limit
Vehicle_Propulsion_Code
Lighting
Pedestrian1_Movement
Pedestrian1_Movement
Vehicle_1st_Point_of_Impact
Pedestrian1_Age
Pedestrian1_Location
Vehicle_Type
Pedestrian1_Age
Vehicle_Manoeuvre

21



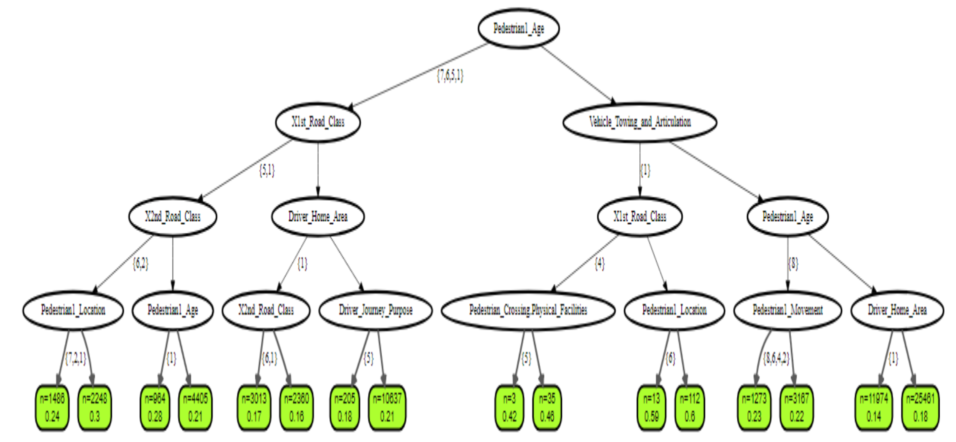
Vehicle_Type
{6,5,3,1}
Area
Vehicle_Towing_and_Articulation
{2}
{1}
X1st_Road_Class
Speed_limit
Speed_limit
Vehicle_1st_Point_of_Impact
{5,1}
{4,1}
{2,1}
{5,2}
Pedestrian1_Age
Lighting
Pavement
Number_of_Pedestrian
Area
Pedestrian1_Age
Speed_limit
Pedestrian1_Location
{7,2}
{5,2}
{1}
{2}
{2}
{7,1}
{1}
{7,6}
n=557 0.29
n=1539 0.39
n=4156 0.21
n=464 0.41
n=1276 0.26
n=477 0.3
n=48060 0.17
n=1574 0.18
n=83 0.74
n=31 0.55
n=59 0.53
n=140 0.5
n=206 0.54
n=4387 0.22
n=1525 0.19
n=2822 0.21

22



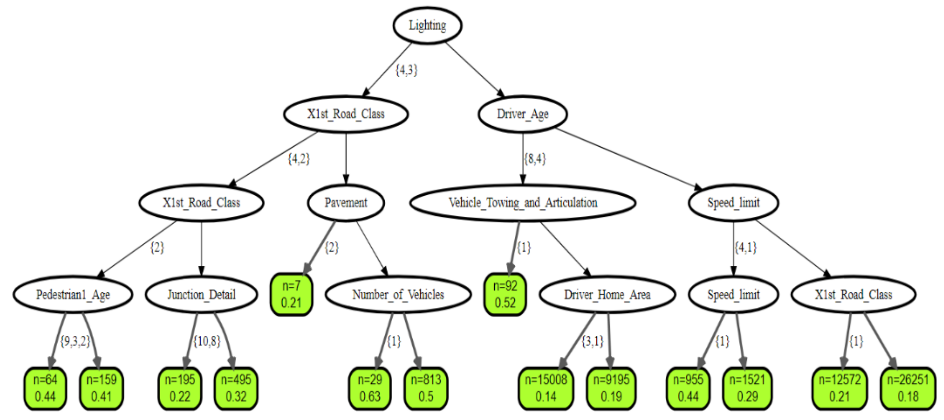
Speed_limit
X2nd_Road_Class
Driver_Journey_Purpose
Driver_Age
Vehicle_Type
Pedestrian1_Age
Pedestrian1_Age
Lighting
Vehicle_1st_Point_of_Impact
Vehicle_Engine_Capacity_CC
Driver_Home_Area
Driver_IMD_Decile
Pedestrian1_Age
Vehicle_Engine_Capacity_CC
Pedestrian1_Movement

23



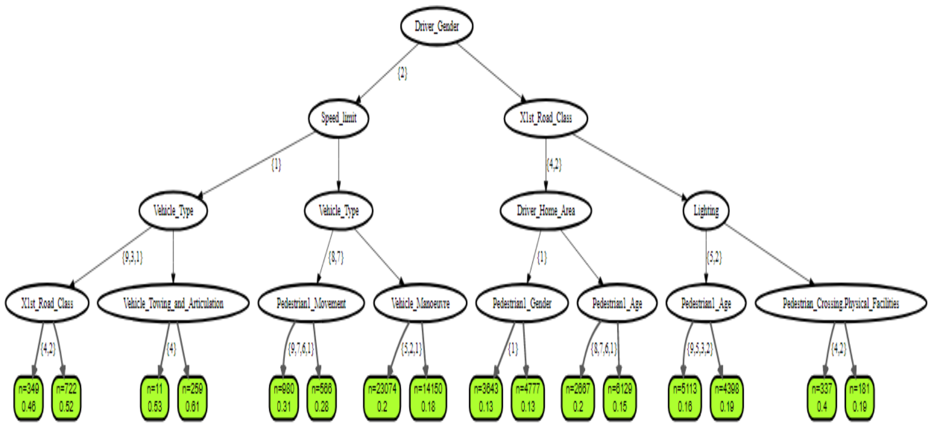
Pedestrian1_Age
X1st_Road_Class
Vehicle_Towing_and_Articulation
X2nd_Road_Class
Driver_Home_Area
X1st_Road_Class
Pedestrian1_Age
Pedestrian1_Location
Pedestrian1_Age
X2nd_Road_Class
Driver_Journey_Purpose
Pedestrian_Crossing_Physical_Facilities
Pedestrian1_Location
Pedestrian1_Movement
Driver_Home_Area

24

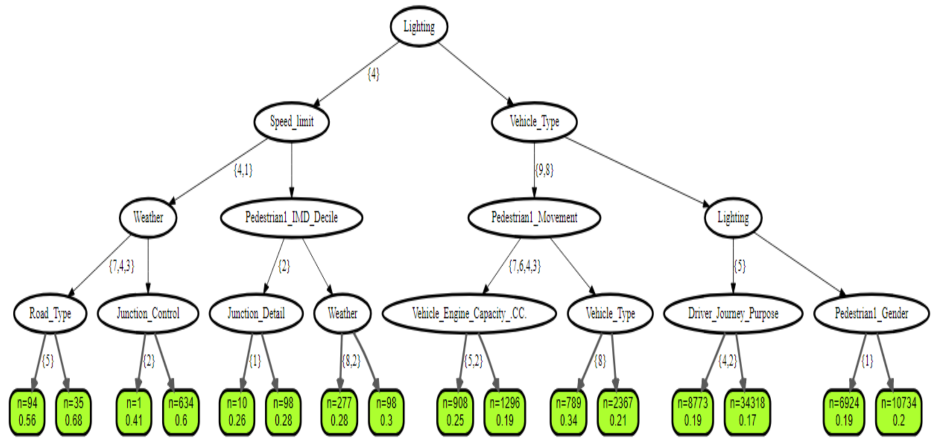


Lighting
{4,3}
X1st_Road_Class
Driver_Age
{4,2}
{8,4}
X1st_Road_Class
Pavement
Vehicle_Towing_and_Articulation
Speed_limit
{2}
{2}
{1}
{4,1}
Pedestrian1_Age
Junction_Detail
n=7 0.21
Number_of_Vehicles
n=92 0.52
Driver_Home_Area
Speed_limit
X1st_Road_Class
{9,3,2}
{10,8}
{1}
{3,1}
{1}
{1}
n=64 0.44
n=159 0.41
n=195 0.22
n=495 0.32
n=29 0.63
n=813 0.5
n=15008 0.14
n=9195 0.19
n=955 0.44
n=1521 0.29
n=12572 0.21
n=26251 0.18

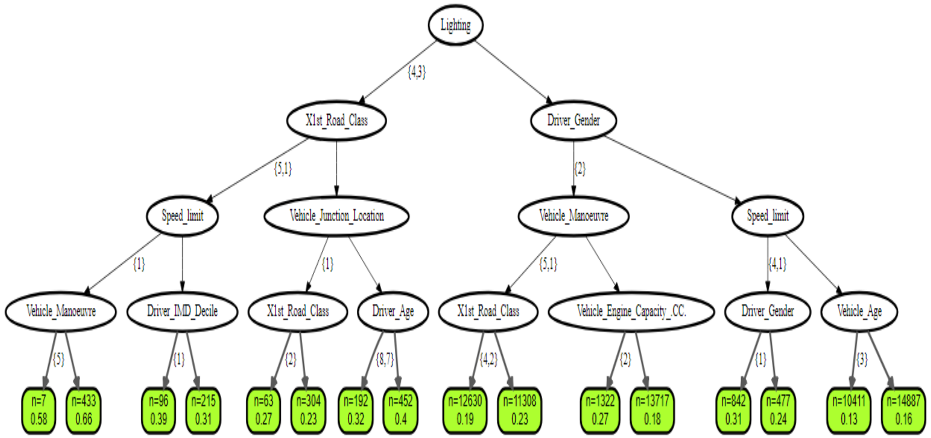
25


Driver_Gender
{2}
Speed_limit
X1st_Road_Class
{1}
Vehicle_Type
Vehicle_Type
{4,2}
Driver_Home_Area
Lighting
{9,3,1}
{8,7}
{1}
{5,2}
X1st_Road_Class
Vehicle_Towing_and_Articulation
Pedestrian1_Movement
Vehicle_Manoeuvre
Pedestrian1_Gender
Pedestrian1_Age
Pedestrian1_Age
Pedestrian_Crossing.Physical_Facilities
{4,2}
{4}
{9,7,6,1}
{5,2,1}
{1}
{8,7,6,1}
{9,5,3,2}
{4,2}
n=349 0.46
n=722 0.52
n=11 0.53
n=259 0.61
n=980 0.31
n=566 0.28
n=23074 0.2
n=14150 0.18
n=3643 0.13
n=4777 0.13
n=2667 0.2
n=6129 0.15
n=5113 0.16
n=4398 0.19
n=337 0.4
n=181 0.19

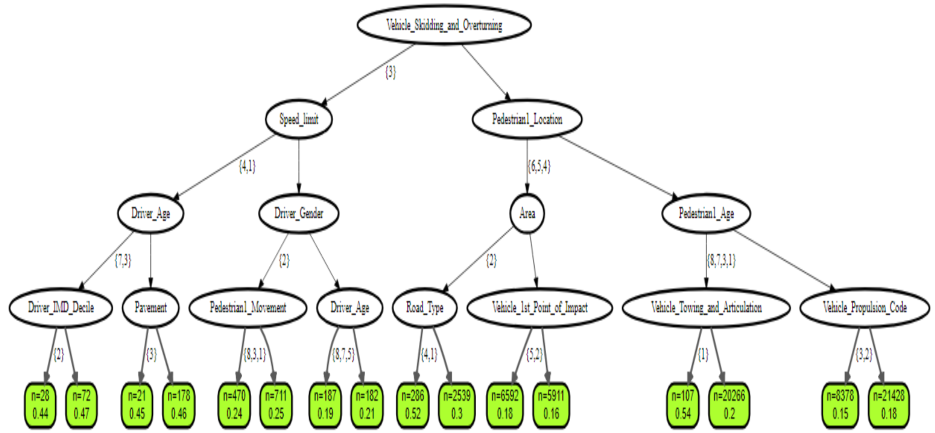
26


Lighting
{4}
Speed_limit
Vehicle_Type
{4,1}
{9,8}
Weather
Pedestrian1_IMD_Decile
Pedestrian1_Movement
Lighting
{7,4,3}
{2}
{7,6,4,3}
{5}
Road_Type
Junction_Control
Junction_Detail
Weather
Vehicle_Engine_Capacity_.CC.
Vehicle_Type
Driver_Journey_Purpose
Pedestrian1_Gender
{5}
{2}
{1}
{8,2}
{5,2}
{8}
{4,2}
{1}
n=94 0.56
n=35 0.68
n=1 0.41
n=634 0.6
n=10 0.26
n=98 0.28
n=277 0.28
n=98 0.3
n=908 0.25
n=1296 0.19
n=789 0.34
n=2367 0.21
n=8773 0.19
n=34318 0.17
n=6924 0.19
n=10734 0.2

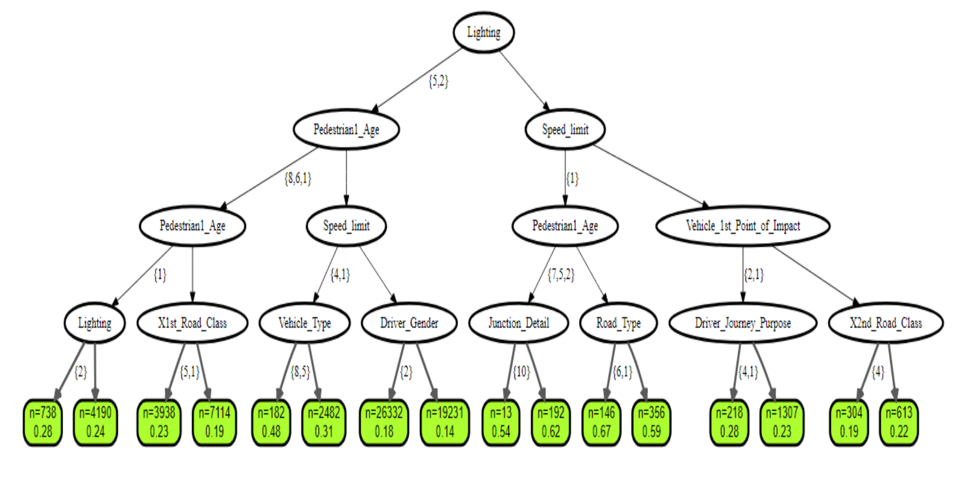
27


Lighting
{4,3}
X1st_Road_Class
Driver_Gender
{5,1}
{2}
Speed_limit
Vehicle_Junction_Location
Vehicle_Manoeuvre
Speed_limit
{1}
{1}
{5,1}
{4,1}
Vehicle_Manoeuvre
Driver_IMD_Decile
X1st_Road_Class
Driver_Age
X1st_Road_Class
Vehicle_Engine_Capacity_.CC.
Driver_Gender
Vehicle_Age
{5}
{1}
{2}
{8,7}
{4,2}
{2}
{1}
{3}
n=7 0.58
n=433 0.66
n=96 0.39
n=215 0.31
n=63 0.27
n=304 0.23
n=192 0.32
n=452 0.4
n=12630 0.19
n=11308 0.23
n=1322 0.27
n=13717 0.18
n=842 0.31
n=477 0.24
n=10411 0.13
n=14887 0.16

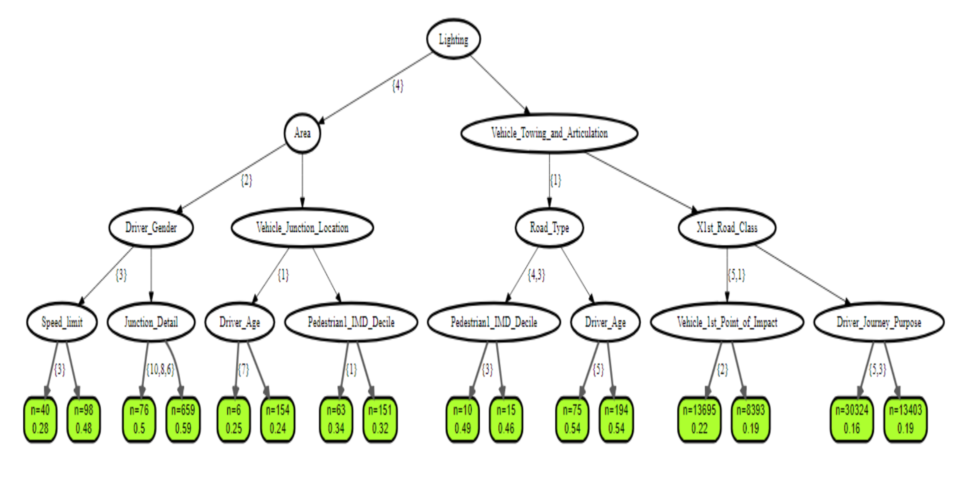
28


Vehicle_Skidding_and_Overturning
{3}
Speed_limit
Pedestrian1_Location
{4,1}
{6,5,4}
Driver_Age
Driver_Gender
Area
Pedestrian1_Age
{7,3}
{2}
{2}
{8,7,3,1}
Driver_IMD_Decile
Pavement
Pedestrian1_Movement
Driver_Age
Road_Type
Vehicle_1st_Point_of_Impact
Vehicle_Towing_and_Articulation
Vehicle_Propulsion_Code
{2}
{3}
{8,3,1}
{8,7,5}
{4,1}
{5,2}
{1}
{3,2}
n=28 0.44
n=72 0.47
n=21 0.45
n=178 0.46
n=470 0.24
n=711 0.25
n=187 0.19
n=182 0.21
n=286 0.52
n=2539 0.3
n=6592 0.18
n=5911 0.16
n=107 0.54
n=20266 0.2
n=8378 0.15
n=21428 0.18

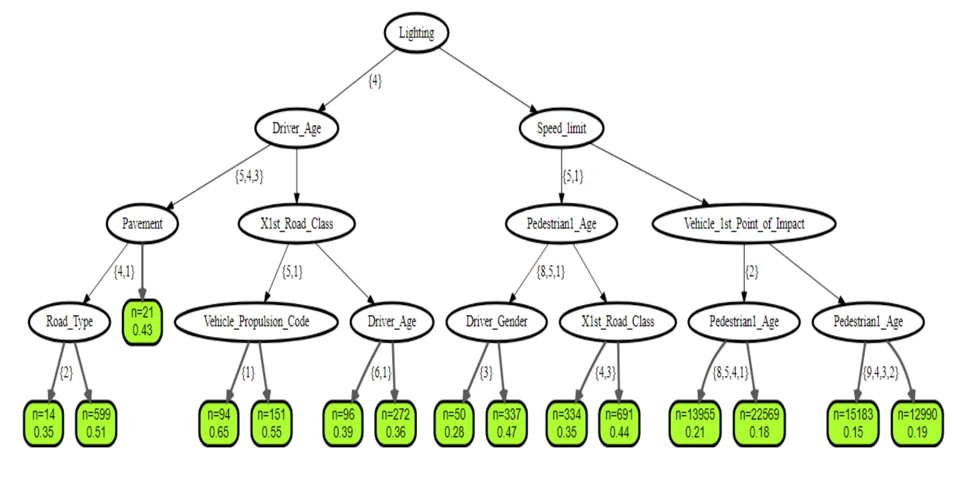
29


Lighting
{5,2}
Pedestrian1_Age
Speed_limit
{8,6,1}
Pedestrian1_Age
Speed_limit
{1}
Pedestrian1_Age
Vehicle_1st_Point_of_Impact
{1}
Lighting
X1st_Road_Class
{4,1}
Vehicle_Type
Driver_Gender
{7,5,2}
Junction_Detail
Road_Type
{2,1}
Driver_Journey_Purpose
X2nd_Road_Class
{2}
{5,1}
{8,5}
{2}
{10}
{6,1}
{4,1}
{4}
n=738 0.28
n=4190 0.24
n=3938 0.23
n=7114 0.19
n=182 0.48
n=2482 0.31
n=26332 0.18
n=19231 0.14
n=13 0.54
n=192 0.62
n=146 0.67
n=356 0.59
n=218 0.28
n=1307 0.23
n=304 0.19
n=613 0.22

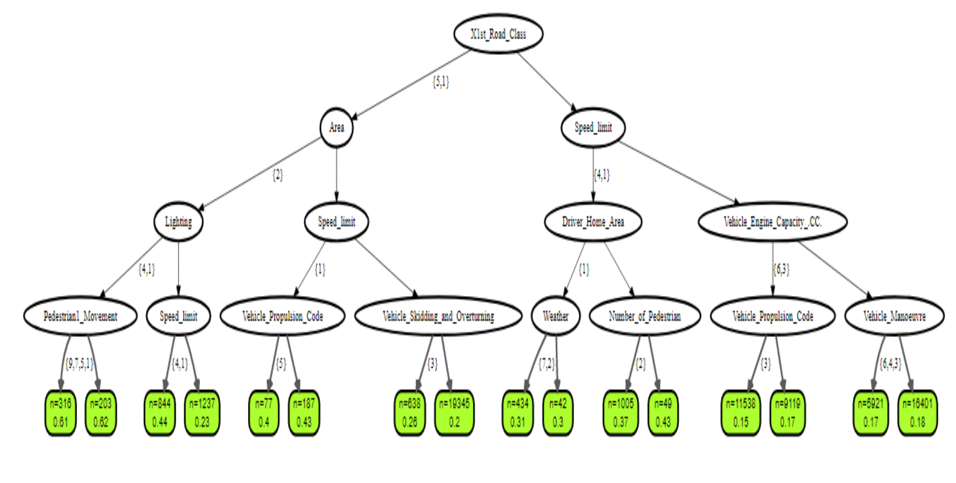
30


Lighting
{4}
Area
Vehicle_Towing_and_Articulation
{2}
Driver_Gender
Vehicle_Junction_Location
{1}
Road_Type
X1st_Road_Class
{3}
Speed_limit
Junction_Detail
{1}
Driver_Age
Pedestrian1_IMD_Decile
{4,3}
Pedestrian1_IMD_Decile
Driver_Age
{5,1}
Vehicle_1st_Point_of_Impact
Driver_Journey_Purpose
{3}
{10,8,6}
{7}
{1}
{3}
{5}
{2}
{5,3}
n=40 0.28
n=98 0.48
n=76 0.5
n=659 0.59
n=6 0.25
n=154 0.24
n=63 0.34
n=151 0.32
n=10 0.49
n=15 0.46
n=75 0.54
n=194 0.54
n=13695 0.22
n=8393 0.19
n=30324 0.16
n=13403 0.19

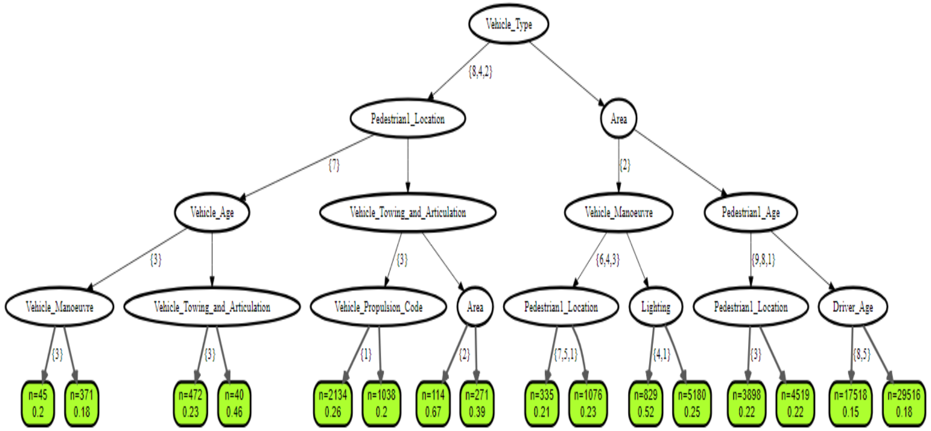
31


Lighting
{4}
Driver_Age
Speed_limit
{5,4,3}
Pavement
X1st_Road_Class
{5,1}
Pedestrian1_Age
Vehicle_1st_Point_of_Impact
{4,1}
Road_Type
n=21 0.43
{5,1}
Vehicle_Propulsion_Code
Driver_Age
{8,5,1}
Driver_Gender
X1st_Road_Class
{2}
Pedestrian1_Age
Pedestrian1_Age
{2}
{1}
{6,1}
{3}
{4,3}
{8,5,4,1}
{9,4,3,2}
n=14 0.35
n=599 0.51
n=94 0.65
n=151 0.55
n=96 0.39
n=272 0.36
n=50 0.28
n=337 0.47
n=334 0.35
n=691 0.44
n=13955 0.21
n=22569 0.18
n=15183 0.15
n=12990 0.19

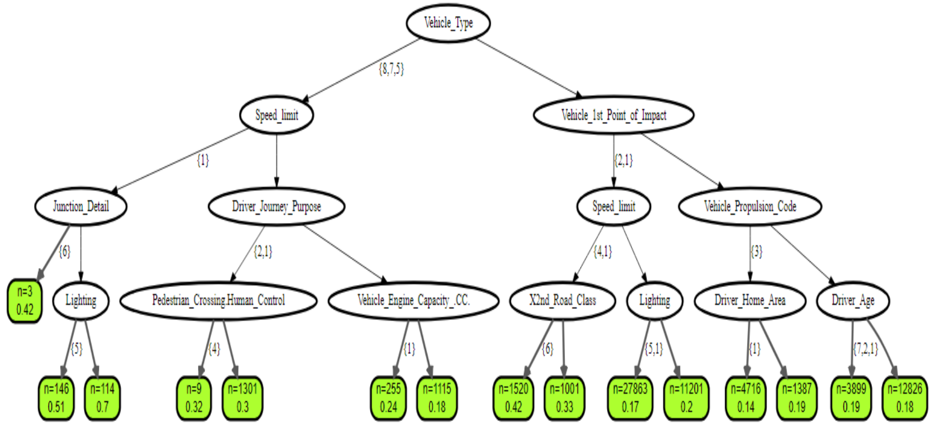
32


X1st_Road_Class
{5,1}
Area
Speed_limit
{2}
Lighting
Speed_limit
{4,1}
Driver_Home_Area
Vehicle_Engine_Capacity_CC.
{4,1}
Pedestrian1_Movement
Speed_limit
{1}
Vehicle_Propulsion_Code
Vehicle_Skidding_and_Overturning
{1}
Weather
Number_of_Pedestrian
{6,3}
Vehicle_Propulsion_Code
Vehicle_Manoeuvre
{9,7,5,1}
{4,1}
{5}
{3}
{7,2}
{2}
{3}
{6,4,3}
n=316 0.61
n=203 0.62
n=844 0.44
n=1237 0.23
n=77 0.4
n=187 0.43
n=638 0.26
n=19345 0.2
n=434 0.31
n=42 0.3
n=1005 0.37
n=49 0.43
n=11538 0.15
n=9119 0.17
n=5921 0.17
n=16401 0.18

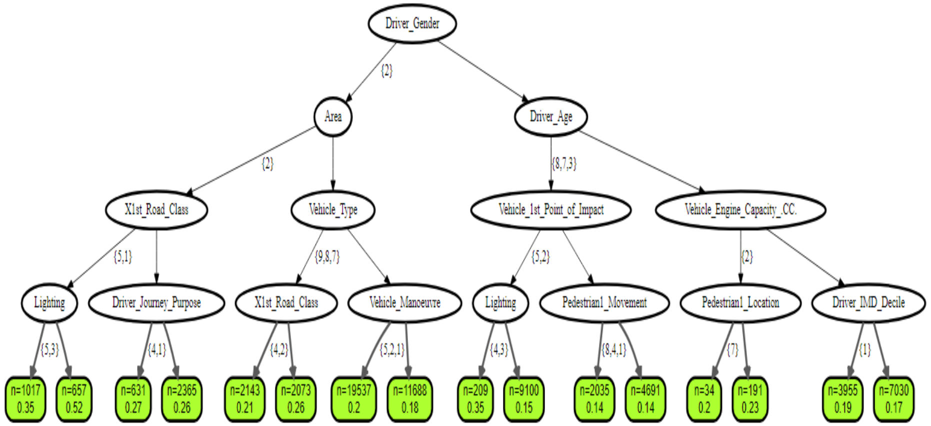
33


Vehicle_Type
{8,4,2}
Pedestrian1_Location
Area
{7}
{2}
Vehicle_Age
Vehicle_Towing_and_Articulation
Vehicle_Manoeuvre
Pedestrian1_Age
{3}
{3}
{6,4,3}
{9,8,1}
Vehicle_Manoeuvre
Vehicle_Towing_and_Articulation
Vehicle_Propulsion_Code
Area
Pedestrian1_Location
Lighting
Pedestrian1_Location
Driver_Age
{3}
{3}
{1}
{2}
{7,5,1}
{4,1}
{3}
{8,5}
n=45 0.2
n=371 0.18
n=472 0.23
n=40 0.46
n=2134 0.26
n=1038 0.2
n=114 0.67
n=271 0.39
n=335 0.21
n=1076 0.23
n=829 0.52
n=5180 0.25
n=3898 0.22
n=4519 0.22
n=17518 0.15
n=29516 0.18

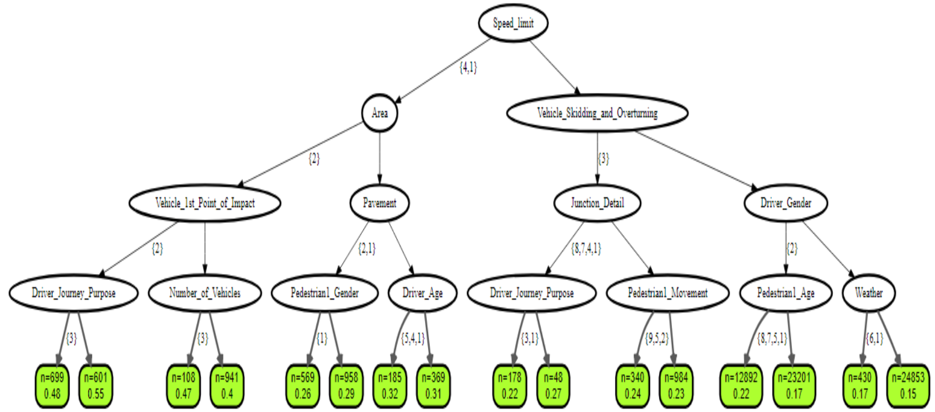
34


Vehicle_Type
{8,7,5}
Speed_limit
Vehicle_1st_Point_of_Impact
{1}
{2,1}
Junction_Detail
Driver_Journey_Purpose
Speed_limit
Vehicle_Propulsion_Code
{6}
{2,1}
{4,1}
{3}
n=3 0.42
Lighting
Pedestrian_Crossing.Human_Control
Vehicle_Engine_Capacity_.CC.
X2nd_Road_Class
Lighting
Driver_Home_Area
Driver_Age
{5}
{4}
{1}
{6}
{5,1}
{1}
{7,2,1}
n=146 0.51
n=114 0.7
n=9 0.32
n=1301 0.3
n=255 0.24
n=1115 0.18
n=1520 0.42
n=1001 0.33
n=27863 0.17
n=11201 0.2
n=4716 0.14
n=1387 0.19
n=3899 0.19
n=12826 0.18

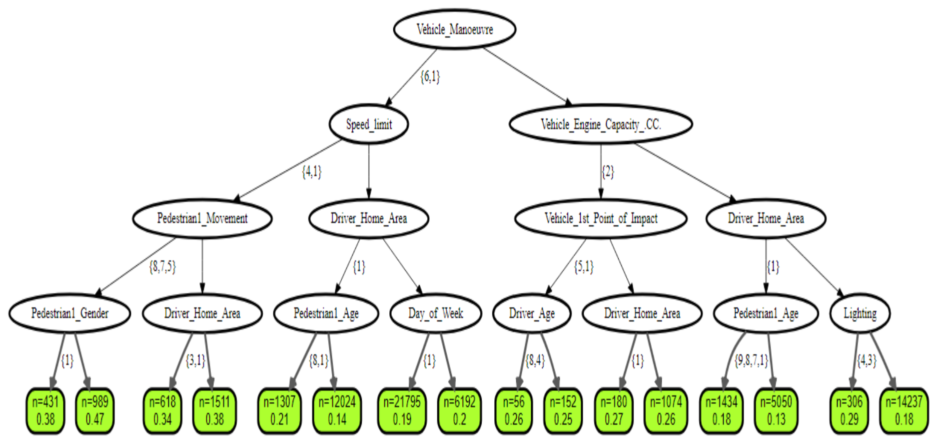
35


Driver_Gender
{2}
Area
Driver_Age
{2}
{8,7,3}
X1st_Road_Class
Vehicle_Type
Vehicle_1st_Point_of_Impact
Vehicle_Engine_Capacity_.CC.
{5,1}
{9,8,7}
{5,2}
{2}
Lighting
Driver_Journey_Purpose
X1st_Road_Class
Vehicle_Manoeuvre
Lighting
Pedestrian1_Movement
Pedestrian1_Location
Driver_IMD_Decile
{5,3}
{4,1}
{4,2}
{5,2,1}
{4,3}
{8,4,1}
{7}
{1}
n=1017 0.35
n=657 0.52
n=631 0.27
n=2365 0.26
n=2143 0.21
n=2073 0.26
n=19537 0.2
n=11688 0.18
n=209 0.35
n=9100 0.15
n=2035 0.14
n=4691 0.14
n=34 0.2
n=191 0.23
n=3955 0.19
n=7030 0.17

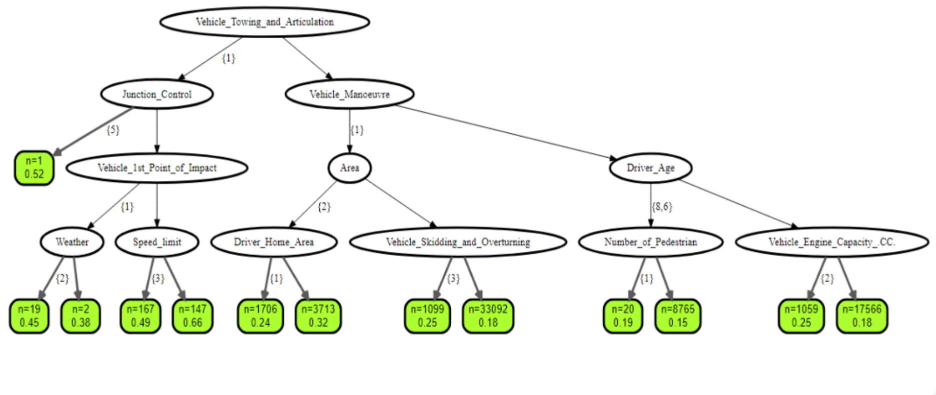
36


Speed_limit
{4,1}
Area
Vehicle_Skidding_and_Overturning
{2}
{3}
Vehicle_1st_Point_of_Impact
Pavement
Junction_Detail
Driver_Gender
{2}
{2,1}
{8,7,4,1}
{2}
Driver_Journey_Purpose
Number_of_Vehicles
Pedestrian1_Gender
Driver_Age
Driver_Journey_Purpose
Pedestrian1_Movement
Pedestrian1_Age
Weather
{3}
{3}
{1}
{5,4,1}
{3,1}
{9,5,2}
{8,7,5,1}
{6,1}
n=699 0.48
n=601 0.55
n=108 0.47
n=941 0.4
n=569 0.26
n=958 0.29
n=185 0.32
n=369 0.31
n=178 0.22
n=48 0.27
n=340 0.24
n=984 0.23
n=12892 0.22
n=23201 0.17
n=430 0.17
n=24853 0.15

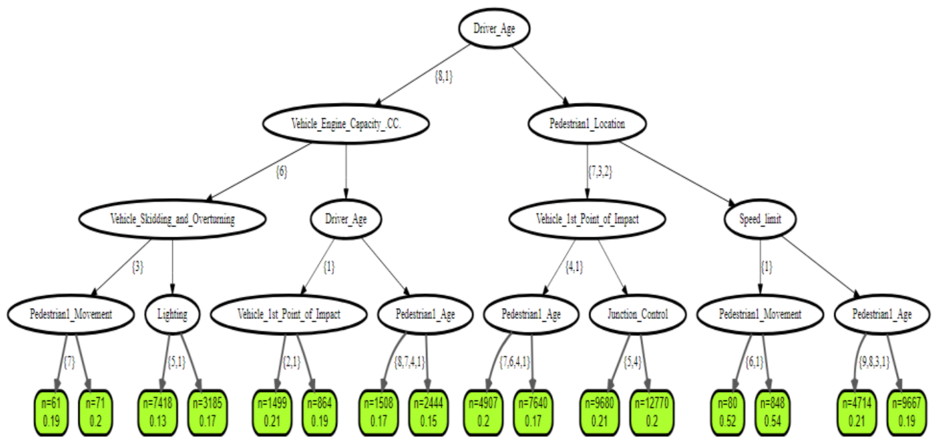
37


Vehicle_Manoeuvre
{6,1}
Speed_limit
Vehicle_Engine_Capacity_CC.
{4,1}
Pedestrian1_Movement
Driver_Home_Area
{2}
Vehicle_1st_Point_of_Impact
Driver_Home_Area
{8,7,5}
Pedestrian1_Gender
Driver_Home_Area
{1}
Pedestrian1_Age
Day_of_Week
{5,1}
Driver_Age
Driver_Home_Area
{1}
Pedestrian1_Age
Lighting
{1}
n=431 0.38
n=989 0.47
{3,1}
n=618 0.34
n=1511 0.38
{8,1}
n=1307 0.21
n=12024 0.14
{1}
n=21795 0.19
n=6192 0.2
{8,4}
n=56 0.26
n=152 0.25
{1}
n=180 0.27
n=1074 0.26
{9,8,7,1}
n=1434 0.18
n=5050 0.13
{4,3}
n=306 0.29
n=14237 0.18

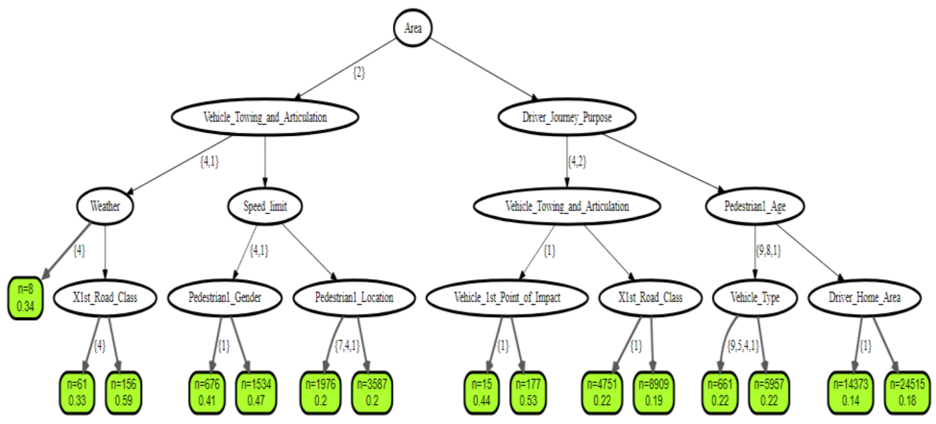
38


Vehicle_Towing_and_Articulation
{1}
Junction_Control
Vehicle_Manoeuvre
{5}
n=1 0.52
Vehicle_1st_Point_of_Impact
{1}
Area
Driver_Age
{1}
Weather
Speed_limit
{2}
Driver_Home_Area
Vehicle_Skidding_and_Overturning
{8,6}
Number_of_Pedestrian
Vehicle_Engine_Capacity_CC.
{2}
n=19 0.45
n=2 0.38
{3}
n=167 0.49
n=147 0.66
{1}
n=1706 0.24
n=3713 0.32
{3}
n=1099 0.25
n=33092 0.18
{1}
n=20 0.19
n=8765 0.15
{2}
n=1059 0.25
n=17566 0.18

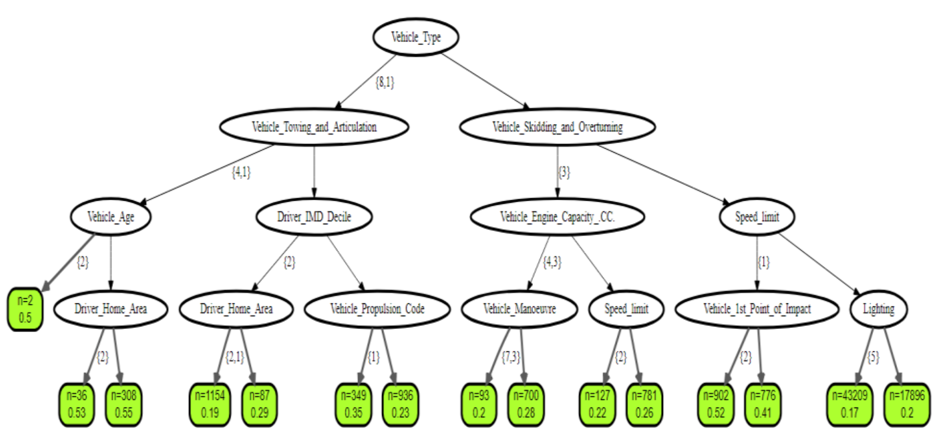
39


Driver_Age
{8,1}
Vehicle_Engine_Capacity_CC.
Pedestrian1_Location
{6}
Vehicle_Skidding_and_Overturning
Driver_Age
{7,3,2}
Vehicle_1st_Point_of_Impact
Speed_limit
{3}
Pedestrian1_Movement
Lighting
{1}
Vehicle_1st_Point_of_Impact
Pedestrian1_Age
{4,1}
Pedestrian1_Age
Junction_Control
{1}
Pedestrian1_Movement
Pedestrian1_Age
{7}
n=61 0.19
n=71 0.2
{5,1}
n=7418 0.13
n=3185 0.17
{2,1}
n=1499 0.21
n=864 0.19
{8,7,4,1}
n=1508 0.17
n=2444 0.15
{7,6,4,1}
n=4907 0.2
n=7640 0.17
{5,4}
n=9680 0.21
n=12770 0.2
{6,1}
n=80 0.52
n=848 0.54
{9,8,3,1}
n=4714 0.21
n=9667 0.19

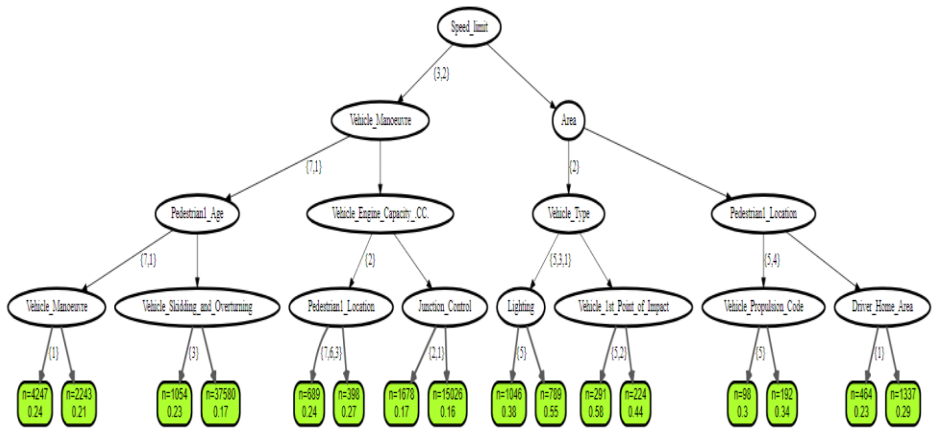
40


Area
{2}
Vehicle_Towing_and_Articulation
Driver_Journey_Purpose
{4,1}
Weather
Speed_limit
{4,2}
Vehicle_Towing_and_Articulation
Pedestrian1_Age
{4}
n=8 0.34
X1st_Road_Class
{4,1}
Pedestrian1_Gender
Pedestrian1_Location
{1}
Vehicle_1st_Point_of_Impact
X1st_Road_Class
{9,8,1}
Vehicle_Type
Driver_Home_Area
{4}
n=61 0.33
n=156 0.59
{1}
n=676 0.41
n=1534 0.47
{7,4,1}
n=1976 0.2
n=3587 0.2
{1}
n=15 0.44
n=177 0.53
{1}
n=4751 0.22
n=8909 0.19
{9,5,4,1}
n=661 0.22
n=5967 0.22
{1}
n=14373 0.14
n=24515 0.18

41


Vehicle_Type
{8,1}
Vehicle_Towing_and_Articulation
Vehicle_Skidding_and_Overturning
{4,1}
{3}
Vehicle_Age
Driver_IMD_Decile
Vehicle_Engine_Capacity_CC.
Speed_limit
{2}
{2}
{4,3}
{1}
n=2
0.5
Driver_Home_Area
Driver_Home_Area
Vehicle_Propulsion_Code
Vehicle_Manoeuvre
Speed_limit
Vehicle_1st_Point_of_Impact
Lighting
{2}
{2,1}
{1}
{7,3}
{2}
{2}
{5}
n=36
0.53
n=308
0.55
n=1154
0.19
n=87
0.29
n=349
0.35
n=936
0.23
n=93
0.2
n=700
0.28
n=127
0.22
n=781
0.26
n=902
0.52
n=776
0.41
n=43209
0.17
n=17896
0.2

42


Speed_limit
{3,2}
Vehicle_Manoeuvre
Area
{7,1}
{2}
Pedestrian1_Age
Vehicle_Engine_Capacity_CC.
Vehicle_Type
Pedestrian1_Location
{7,1}
{2}
{5,3,1}
{5,4}
Vehicle_Manoeuvre
Vehicle_Skidding_and_Overturning
Pedestrian1_Location
Junction_Control
Lighting
Vehicle_1st_Point_of_Impact
Vehicle_Propulsion_Code
Driver_Home_Area
{1}
{3}
{7,6,3}
{2,1}
{5}
{5,2}
{5}
{1}
n=4247
0.24
n=2243
0.21
n=1054
0.23
n=37580
0.17
n=689
0.24
n=398
0.27
n=1678
0.17
n=15026
0.16
n=1046
0.38
n=789
0.55
n=291
0.58
n=224
0.44
n=98
0.3
n=192
0.34
n=464
0.23
n=1337
0.29

Association rules

Rules with fatal as consequent

Table 145 – Association rules with vehicles characteristics as first antecedent and fatal crashes as consequent, Great Britain.

ID rule	Rules with vehicles characteristics as first antecedent and fatal pedestrian crash as consequent Antecents	S %	C %	Lift	LIC
1	Vehicle Towing and Articulation=Articulated vehicle	0.14	28.87	14.24	n.a.
2	Vehicle Towing and Articulation=Articulated vehicle & 1st Road Class=A	0.10	34.33	16.93	1.19
3	Vehicle Towing and Articulation=Articulated vehicle & 1st Road Class=A & Vehicle Type=Truck	0.10	36.17	17.84	1.05
4	Vehicle Towing and Articulation=Articulated vehicle & Vehicle Type=Truck	0.14	31.27	15.42	1.08
5	Vehicle Towing and Articulation=Articulated vehicle & Vehicle Type=Truck & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.10	33.50	16.52	1.07
6	Vehicle Towing and Articulation=Articulated vehicle & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.10	31.11	15.34	1.08
7	Vehicle Type=Truck	0.30	13.64	6.73	n.a.
8	Vehicle Type=Truck & Speed limit≥50	0.12	40.21	19.83	2.95
9	Vehicle Type=Truck & Road Type=Dual carriageway	0.10	29.96	14.77	2.20
10	Vehicle Type=Truck & Area=Rural	0.13	24.08	11.87	1.77
11	Vehicle Type=Truck & Area=Rural & Junction Detail=Not at junction	0.10	25.65	12.65	1.07
12	Vehicle Type=Truck & Area=Rural & Junction Control=Not at junction or within 20 metres	0.10	25.56	12.60	1.06
13	Vehicle Type=Truck & 1st Road Class=A	0.20	20.95	10.33	1.54
14	Vehicle Type=Truck & 1st Road Class=A & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.12	23.14	11.41	1.10
15	Vehicle Type=Truck & 1st Road Class=A & Junction Detail=Not at junction	0.11	22.77	11.23	1.09
16	Vehicle Type=Truck & 1st Road Class=A & Junction Control=Not at junction or within 20 metres	0.11	22.36	11.02	1.07
17	Vehicle Type=Truck & Junction Detail=Not at junction	0.17	14.64	7.22	1.07
18	Vehicle Type=Truck & Junction Control=Not at junction or within 20 metres	0.17	14.34	7.07	1.05
19	Vehicle Skidding and Overturning=Yes	0.21	7.63	3.76	n.a.
20	Vehicle Skidding and Overturning=Yes & 1st Road Class=A	0.13	10.53	5.19	1.38
21	Vehicle Skidding and Overturning=Yes & 1st Road Class=A & Vehicle Manoeuvre=Going ahead	0.12	13.25	6.53	1.26
22	Vehicle Skidding and Overturning=Yes & 1st Road Class=A & Weather=Fine no high winds	0.11	11.94	5.89	1.13
23	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead	0.20	9.49	4.68	1.24
24	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Pavement=Dry	0.14	10.96	5.40	1.15
25	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Weather=Fine no high winds	0.18	10.94	5.39	1.15
26	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.15	10.82	5.33	1.14
27	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Vehicle Location=Not at junction	0.12	10.67	5.26	1.12
28	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Junction Detail=Not at junction	0.12	10.67	5.26	1.12
29	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Junction Control=Not at junction or within 20 metres	0.12	10.65	5.25	1.12

ID rule	Rules with vehicles characteristics as first antecedent and fatal pedestrian crash as consequent Antecents	S %	C %	Lift	LIC
30	Vehicle Skidding and Overturning=Yes & Vehicle Manoeuvre=Going ahead & Vehicle Type=Car	0.12	10.02	4.94	1.06
31	Vehicle Skidding and Overturning=Yes & Vehicle Location=Not at junction	0.12	9.22	4.55	1.21
32	Vehicle Skidding and Overturning=Yes & Junction Detail=Not at junction	0.12	9.22	4.55	1.21
33	Vehicle Skidding and Overturning=Yes & Vehicle Location=Not at junction & Weather=Fine no high winds	0.11	10.82	5.34	1.17
34	Vehicle Skidding and Overturning=Yes & Junction Detail=Not at junction & Weather=Fine no high winds	0.11	10.82	5.34	1.17
35	Vehicle Skidding and Overturning=Yes & Junction Control=Not at junction or within 20 metres	0.12	9.19	4.53	1.21
36	Vehicle Skidding and Overturning=Yes & Junction Control=Not at junction or within 20 metres & Weather=Fine no high winds	0.11	10.78	5.31	1.17
37	Vehicle Skidding and Overturning=Yes & Weather=Fine no high winds	0.19	8.92	4.40	1.17
38	Vehicle Skidding and Overturning=Yes & Weather=Fine no high winds & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.14	10.40	5.13	1.17
39	Vehicle Skidding and Overturning=Yes & Weather=Fine no high winds & Vehicle Type=Car	0.12	9.84	4.85	1.10
40	Vehicle Skidding and Overturning=Yes & Pavement=Dry	0.15	8.85	4.36	1.16
41	Vehicle Skidding and Overturning=Yes & Pavement=Dry & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.11	10.27	5.07	1.16
42	Vehicle Skidding and Overturning=Yes & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.15	8.80	4.34	1.15
43	Vehicle Skidding and Overturning=Yes & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres & Day of Week=Weekday	0.12	9.29	4.58	1.06
44	Vehicle Skidding and Overturning=Yes & Vehicle Type=Car	0.13	8.22	4.05	1.08
45	Vehicle Engine Capacity ≥3000	0.35	6.89	3.40	n.a.
46	Vehicle Engine Capacity ≥3000 & Speed limit≥50	0.10	39.53	19.49	5.74
47	Vehicle Engine Capacity ≥3000 & Vehicle Towing and Articulation=Articulated vehicle	0.12	32.92	16.23	4.78
48	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck	0.22	22.30	11.00	3.24
49	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Area=Rural	0.10	39.66	19.55	1.78
50	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Vehicle 1st Point of Impact=Front	0.15	32.47	16.01	1.46
51	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & 1st Road Class=A	0.15	27.39	13.51	1.23
52	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Junction Detail=Not at junction	0.13	26.71	13.17	1.20
53	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Vehicle Location=Not at junction	0.13	26.63	13.13	1.19
54	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Junction Control=Not at junction or within 20 metres	0.13	26.55	13.09	1.19
55	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Vehicle Manoeuvre=Going ahead	0.13	24.26	11.96	1.09
56	Vehicle Engine Capacity ≥3000 & Vehicle Type=Truck & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.15	23.62	11.65	1.06
57	Vehicle Engine Capacity ≥3000 & Area=Rural	0.11	20.16	9.94	2.92
58	Vehicle Engine Capacity ≥3000 & Area=Rural & Driver Gender=M	0.11	21.45	10.58	1.06
59	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front	0.24	9.48	4.67	1.37
60	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & Junction Detail=Not at junction	0.13	11.83	5.83	1.25
61	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & Vehicle Location=Not at junction	0.13	11.80	5.82	1.24
62	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & Junction Control=Not at junction or within 20 metres	0.13	11.75	5.79	1.24
63	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.15	11.46	5.65	1.21
64	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & 1st Road Class=A	0.15	11.43	5.63	1.21
65	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & Driver Journey Purpose=Journey as part of work	0.21	11.23	5.54	1.18
66	Vehicle Engine Capacity ≥3000 & Vehicle 1st Point of Impact=Front & Vehicle Manoeuvre=Going ahead	0.16	10.39	5.12	1.10

ID rule	Rules with vehicles characteristics as first antecedent and fatal pedestrian crash as consequent Antecents	S %	C %	Lift	LIC
67	Vehicle Engine Capacity ≥3000 & 1st Road Class=A	0.22	9.14	4.51	1.33
68	Vehicle Engine Capacity ≥3000 & 1st Road Class=A & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.12	11.80	5.82	1.29
69	Vehicle Engine Capacity ≥3000 & 1st Road Class=A & Junction Detail=Not at junction	0.11	11.51	5.67	1.26
70	Vehicle Engine Capacity ≥3000 & 1st Road Class=A & Vehicle Location=Not at junction	0.11	11.49	5.67	1.26
71	Vehicle Engine Capacity ≥3000 & 1st Road Class=A & Junction Control=Not at junction or within 20 metres	0.11	11.35	5.60	1.24
72	Vehicle Engine Capacity ≥3000 & 1st Road Class=A & Vehicle Manoeuvre=Going ahead	0.15	10.65	5.25	1.16
73	Vehicle Engine Capacity ≥3000 & 1st Road Class=A & Driver Journey Purpose=Journey as part of work	0.20	10.43	5.15	1.14
74	Vehicle Engine Capacity ≥3000 & Driver Journey Purpose=Journey as part of work	0.31	8.17	4.03	1.19
75	Vehicle Engine Capacity ≥3000 & Driver Journey Purpose=Journey as part of work & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.19	9.33	4.60	1.14
76	Vehicle Engine Capacity ≥3000 & Driver Journey Purpose=Journey as part of work & Vehicle Manoeuvre=Going ahead	0.19	9.28	4.58	1.14
77	Vehicle Engine Capacity ≥3000 & Driver Journey Purpose=Journey as part of work & Junction Detail=Not at junction	0.16	9.09	4.48	1.11
78	Vehicle Engine Capacity ≥3000 & Driver Journey Purpose=Journey as part of work & Vehicle Location=Not at junction	0.16	9.08	4.47	1.11
79	Vehicle Engine Capacity ≥3000 & Driver Journey Purpose=Journey as part of work & Junction Control=Not at junction or within 20 metres	0.16	9.06	4.47	1.11
80	Vehicle Engine Capacity ≥3000 & Vehicle Manoeuvre=Going ahead	0.21	7.94	3.92	1.15
81	Vehicle Engine Capacity ≥3000 & Vehicle Manoeuvre=Going ahead & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.15	10.07	4.97	1.27
82	Vehicle Engine Capacity ≥3000 & Vehicle Manoeuvre=Going ahead & Junction Detail=Not at junction	0.13	9.28	4.57	1.17
83	Vehicle Engine Capacity ≥3000 & Vehicle Manoeuvre=Going ahead & Vehicle Location=Not at junction	0.13	9.27	4.57	1.17
84	Vehicle Engine Capacity ≥3000 & Vehicle Manoeuvre=Going ahead & Junction Control=Not at junction or within 20 metres	0.13	9.22	4.55	1.16
85	Vehicle Engine Capacity ≥3000 & Vehicle Manoeuvre=Going ahead & Driver Gender=M	0.20	8.40	4.14	1.06
86	Vehicle Engine Capacity ≥3000 & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.21	7.72	3.81	1.12
87	Vehicle Engine Capacity ≥3000 & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres & Driver Gender=M	0.21	8.43	4.16	1.09
88	Vehicle Engine Capacity ≥3000 & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres & Junction Detail=Not at junction	0.13	8.22	4.05	1.06
89	Vehicle Engine Capacity ≥3000 & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres & Vehicle Location=Not at junction	0.13	8.21	4.05	1.06
90	Vehicle Engine Capacity ≥3000 & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres & Junction Control=Not at junction or within 20 metres	0.13	8.18	4.03	1.06
91	Vehicle Engine Capacity ≥3000 & Junction Detail=Not at junction	0.17	7.67	3.78	1.11
92	Vehicle Engine Capacity ≥3000 & Junction Detail=Not at junction & Driver Gender=M	0.17	8.34	4.11	1.09
93	Vehicle Engine Capacity ≥3000 & Vehicle Location=Not at junction	0.17	7.66	3.78	1.11
94	Vehicle Engine Capacity ≥3000 & Vehicle Location=Not at junction & Driver Gender=M	0.17	8.33	4.11	1.09
95	Vehicle Engine Capacity ≥3000 & Junction Control=Not at junction or within 20 metres	0.17	7.56	3.73	1.10
96	Vehicle Engine Capacity ≥3000 & Junction Control=Not at junction or within 20 metres & Driver Gender=M	0.17	8.28	4.08	1.10
97	Vehicle Engine Capacity ≥3000 & Driver Gender=M	0.33	7.33	3.61	1.06

Table 146 – Association rules with environmental characteristics as first antecedent and fatal crashes as consequent, Great Britain..

ID rule	Rules with environmental characteristics as first antecedent and fatal crashes Antecents	S %	C %	Lift	LIC
98	Lighting=Darkness - no lighting	0.33	17.80	8.78	n.a.
99	Lighting=Darkness - no lighting & Road Type=Dual carriageway	0.12	38.92	19.19	2.19
100	Lighting=Darkness - no lighting & Road Type=Dual carriageway & Speed limit≥50	0.11	42.31	20.86	1.09
101	Lighting=Darkness - no lighting & Road Type=Dual carriageway & Junction Detail=Not at junction	0.11	42.11	20.76	1.08
102	Lighting=Darkness - no lighting & Road Type=Dual carriageway & Junction Control=Not at junction or within 20 metres	0.11	42.11	20.76	1.08
103	Lighting=Darkness - no lighting & Speed limit≥50	0.29	30.06	14.82	1.69
104	Lighting=Darkness - no lighting & Speed limit≥50 & 1st Road Class=A	0.21	37.14	18.31	1.24
105	Lighting=Darkness - no lighting & 1st Road Class=A	0.23	29.57	14.58	1.66
106	Lighting=Darkness - no lighting & 1st Road Class=A & Junction Detail=Not at junction	0.20	32.62	16.08	1.10
107	Lighting=Darkness - no lighting & 1st Road Class=A & Junction Control=Not at junction or within 20 metres	0.20	32.46	16.01	1.10
108	Lighting=Darkness - no lighting & Day of Week=Weekend	0.13	21.98	10.84	1.23
109	Lighting=Darkness - no lighting & Day of Week=Weekend & Junction Detail=Not at junction	0.12	24.77	12.22	1.13
110	Lighting=Darkness - no lighting & Day of Week=Weekend & Junction Control=Not at junction or within 20 metres	0.12	24.70	12.18	1.12
111	Lighting=Darkness - no lighting & Junction Detail=Not at junction	0.29	21.13	10.42	1.19
112	Lighting=Darkness - no lighting & Junction Control=Not at junction or within 20 metres	0.29	21.00	10.35	1.18

Table 147 – Association rules with roadway characteristics as first antecedent and fatal crashes as consequent, Great Britain..

ID rule	Rules with roadway characteristics as first antecedent and fatal crashes as consequent Antecents	S %	C %	Lift	LIC
113	Speed limit≥50	0.51	16.74	8.25	n.a.
114	Speed limit≥50 & Road Type=Dual carriageway	0.23	27.34	13.48	1.63
115	Speed limit≥50 & 1st Road Class=A	0.35	21.86	10.78	1.31
116	Speed limit≥50 & 1st Road Class=A & Day of Week=Weekend	0.12	24.02	11.85	1.10
117	Speed limit≥50 & Day of Week=Weekend	0.16	18.41	9.08	1.10
118	Area=Rural	0.68	5.71	2.82	n.a.
119	Area=Rural & Lighting=Darkness - no lighting	0.30	23.02	11.35	4.03
120	Area=Rural & Lighting=Darkness - no lighting & Road Type=Dual carriageway	0.10	40.96	20.20	1.78
121	Area=Rural & Lighting=Darkness - no lighting & 1st Road Class=A	0.21	33.72	16.63	1.46
122	Area=Rural & Lighting=Darkness - no lighting & Speed limit≥50	0.27	29.92	14.75	1.30
123	Area=Rural & Lighting=Darkness - no lighting & Day of Week=Weekend	0.12	26.67	13.15	1.16
124	Area=Rural & Lighting=Darkness - no lighting & Junction Detail=Not at junction	0.27	24.55	12.11	1.07
125	Area=Rural & Lighting=Darkness - no lighting & Junction Control=Not at junction or within 20 metres	0.27	24.55	12.11	1.07
126	Area=Rural & Road Type=Dual carriageway	0.18	22.57	11.13	3.95
127	Area=Rural & Road Type=Dual carriageway & Speed limit≥50	0.17	31.10	15.33	1.38

ID rule	Rules with roadway characteristics as first antecedent and fatal crashes as consequent Antecents	S %	C %	Lift	LIC
128	Area=Rural & Road Type=Dual carriageway & Junction Detail=Not at junction	0.15	25.44	12.54	1.13
129	Area=Rural & Road Type=Dual carriageway & Junction Control=Not at junction or within 20 metres	0.15	25.37	12.51	1.12
130	Area=Rural & Road Type=Dual carriageway & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.17	25.34	12.49	1.12
131	Area=Rural & Speed limit≥50	0.44	17.42	8.59	3.05
132	Area=Rural & Speed limit≥50 & 1st Road Class=A	0.30	23.80	11.73	1.37
133	Area=Rural & Speed limit≥50 & Pavement=Wet or damp	0.14	20.82	10.27	1.20
134	Area=Rural & 1st Road Class=A	0.42	11.43	5.64	2.00
135	Area=Rural & 1st Road Class=A & Day of Week=Weekend	0.15	16.14	7.96	1.41
136	Area=Rural & 1st Road Class=A & Pavement=Wet or damp	0.14	14.24	7.02	1.25
137	Area=Rural & 1st Road Class=A & Junction Detail=Not at junction	0.32	13.46	6.64	1.18
138	Area=Rural & 1st Road Class=A & Junction Control=Not at junction or within 20 metres	0.32	13.35	6.58	1.17
139	Area=Rural & 1st Road Class=A & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.37	12.81	6.32	1.12
140	Area=Rural & Number of Vehicles=2	0.10	10.15	5.00	1.78
141	Area=Rural & Day of Week=Weekend	0.22	8.04	3.96	1.41
142	Area=Rural & Day of Week=Weekend & Junction Detail=Not at junction	0.18	9.24	4.56	1.15
143	Area=Rural & Day of Week=Weekend & Junction Control=Not at junction or within 20 metres	0.18	9.19	4.53	1.14
144	Area=Rural & Pavement=Wet or damp	0.20	7.78	3.84	1.36
145	Area=Rural & Pavement=Wet or damp & Junction Detail=Not at junction	0.16	9.58	4.72	1.23
146	Area=Rural & Pavement=Wet or damp & Junction Control=Not at junction or within 20 metres	0.16	9.54	4.70	1.23
147	Area=Rural & Pavement=Wet or damp & Weather=Fine no high winds	0.12	9.39	4.63	1.21
148	Area=Rural & Pavement=Wet or damp & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.19	8.48	4.18	1.09
149	Area=Rural & Junction Detail=Not at junction	0.50	6.34	3.13	1.11
150	Area=Rural & Junction Control=Not at junction or within 20 metres	0.50	6.31	3.11	1.10
151	Speed limit=40	0.19	5.40	2.66	n.a.
152	Speed limit=40 & Road Type=Single carriageway	0.13	6.78	3.34	1.26
153	Speed limit=40 & 1st Road Class=A	0.14	5.73	2.83	1.06
154	Road Type=Dual carriageway	0.44	5.18	2.56	n.a.
155	Road Type=Dual carriageway & Day of Week=Weekend	0.14	6.84	3.37	1.32
156	Road Type=Dual carriageway & Day of Week=Weekend & 1st Road Class=A	0.12	7.82	3.86	1.14
157	Road Type=Dual carriageway & 1st Road Class=A	0.35	5.63	2.77	1.09

Table 148 – Association rules with pedestrian characteristics as first antecedent and fatal crashes as consequent, Great Britain.

ID rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent Antecents	S %	C %	Lift	LIC
158	Pedestrian Age≥75	0.56	7.46	3.68	n.a.
159	Pedestrian Age≥75 & Lighting=Darkness - lights lit	0.15	13.96	6.88	1.87
160	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Vehicle Manoeuvre=Going ahead	0.13	21.53	10.62	1.54
161	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Vehicle 1st Point of Impact=Front	0.12	16.94	8.35	1.21
162	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Weather=Fine no high winds	0.12	16.27	8.02	1.17
163	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Driver Gender=M	0.11	15.50	7.64	1.11
164	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Driver Home Area=Urban	0.11	14.72	7.26	1.05
165	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Vehicle Age≥15	0.11	14.68	7.24	1.05
166	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead	0.37	12.30	6.07	1.65
167	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Day of Week=Weekend	0.11	17.78	8.77	1.45
168	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & 1st Road Class=A	0.19	16.30	8.04	1.33
169	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Vehicle 1st Point of Impact=Front	0.30	15.39	7.59	1.25
170	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Vehicle Engine Capacity =1500-2000	0.13	14.96	7.37	1.22
171	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver IMD Decile=Less deprived	0.14	14.76	7.28	1.20
172	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Vehicle Propulsion Code=Heavy oil	0.13	14.66	7.23	1.19
173	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver Gender=M	0.28	14.53	7.16	1.18
174	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Pavement=Wet or damp	0.11	14.14	6.97	1.15
175	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Vehicle Age≥15	0.27	13.97	6.89	1.14
176	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Vehicle Propulsion Code=Petrol	0.18	13.87	6.84	1.13
177	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Junction Detail=T or staggered junction	0.12	13.42	6.62	1.09
178	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Pedestrian Gender=M	0.21	13.26	6.54	1.08
179	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver Home Area=Urban	0.25	13.22	6.52	1.07
180	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Weather=Fine no high winds	0.32	13.05	6.44	1.06
181	Pedestrian Age≥75 & Area=Rural	0.12	11.20	5.52	1.50
182	Pedestrian Age≥75 & Area=Rural & Weather=Fine no high winds	0.11	12.17	6.00	1.09
183	Pedestrian Age≥75 & Area=Rural & Road Type=Single carriageway	0.12	12.17	6.00	1.09
184	Pedestrian Age≥75 & 1st Road Class=A	0.26	11.12	5.48	1.49
185	Pedestrian Age≥75 & 1st Road Class=A & Vehicle 1st Point of Impact=Front	0.21	14.46	7.13	1.30
186	Pedestrian Age≥75 & 1st Road Class=A & Junction Detail=Not at junction	0.12	14.00	6.91	1.26
187	Pedestrian Age≥75 & 1st Road Class=A & Vehicle Location=Not at junction	0.11	13.85	6.83	1.25
188	Pedestrian Age≥75 & 1st Road Class=A & Junction Control=Not at junction or within 20 metres	0.12	13.83	6.82	1.24
189	Pedestrian Age≥75 & 1st Road Class=A & Vehicle Propulsion Code=Heavy oil	0.12	13.20	6.51	1.19
190	Pedestrian Age≥75 & 1st Road Class=A & Driver Gender=M	0.20	12.34	6.08	1.11
191	Pedestrian Age≥75 & 1st Road Class=A & Pedestrian Gender=M	0.15	11.98	5.91	1.08
192	Pedestrian Age≥75 & 1st Road Class=A & Driver IMD Decile=More deprived	0.11	11.80	5.82	1.06
193	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front	0.40	10.41	5.13	1.40

ID rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent Antecents	S %	C %	Lift	LIC
194	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Driver Journey Purpose=Journey as part of work	0.10	13.66	6.74	1.31
195	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Vehicle Location=Not at junction	0.18	13.25	6.53	1.27
196	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Junction Detail=Not at junction	0.18	13.25	6.53	1.27
197	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Junction Control=Not at junction or within 20 metres	0.18	13.16	6.49	1.26
198	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Vehicle Propulsion Code=Heavy oil	0.17	12.71	6.27	1.22
199	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Driver Gender=M	0.30	12.22	6.03	1.17
200	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Pavement=Wet or damp	0.12	12.10	5.97	1.16
201	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Weather=Fine no high winds	0.36	11.39	5.62	1.09
202	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Pedestrian Gender=M	0.22	11.37	5.61	1.09
203	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Vehicle Age≥15	0.30	11.18	5.51	1.07
204	Pedestrian Age≥75 & Vehicle 1st Point of Impact=Front & Driver IMD Decile=Less deprived	0.13	11.11	5.48	1.07
205	Pedestrian Age≥75 & Day of Week=Weekend	0.14	9.76	4.81	1.31
206	Pedestrian Age≥75 & Day of Week=Weekend & Driver Gender=M	0.10	11.09	5.47	1.14
207	Pedestrian Age≥75 & Day of Week=Weekend & Weather=Fine no high winds	0.12	10.45	5.15	1.07
208	Pedestrian Age≥75 & Driver Journey Purpose=Journey as part of work	0.16	9.70	4.79	1.30
209	Pedestrian Age≥75 & Driver Journey Purpose=Journey as part of work & Vehicle Propulsion Code=Heavy oil	0.13	11.89	5.86	1.22
210	Pedestrian Age≥75 & Driver Journey Purpose=Journey as part of work & Weather=Fine no high winds	0.14	10.34	5.10	1.07
211	Pedestrian Age≥75 & Driver Journey Purpose=Journey as part of work & Driver Gender=M	0.15	10.24	5.05	1.06
212	Pedestrian Age≥75 & Pavement=Wet or damp	0.15	8.88	4.38	1.19
213	Pedestrian Age≥75 & Pavement=Wet or damp & Weather=Fine no high winds	0.10	12.85	6.34	1.45
214	Pedestrian Age≥75 & Pavement=Wet or damp & Driver Gender=M	0.11	9.95	4.90	1.12
215	Pedestrian Age≥75 & Vehicle Propulsion Code=Heavy oil	0.25	8.82	4.35	1.18
216	Pedestrian Age≥75 & Vehicle Propulsion Code=Heavy oil & Driver Gender=M	0.22	10.07	4.97	1.14
217	Pedestrian Age≥75 & Vehicle Propulsion Code=Heavy oil & Pedestrian Gender=M	0.12	9.32	4.60	1.06
218	Pedestrian Age≥75 & Vehicle Propulsion Code=Heavy oil & Weather=Fine no high winds	0.22	9.28	4.58	1.05
219	Pedestrian Age≥75 & Vehicle Propulsion Code=Heavy oil & Vehicle Location=Not at junction	0.10	9.27	4.57	1.05
220	Pedestrian Age≥75 & Vehicle Propulsion Code=Heavy oil & Junction Detail=Not at junction	0.10	9.27	4.57	1.05
221	Pedestrian Age≥75 & Driver Gender=M	0.43	8.74	4.31	1.17
222	Pedestrian Age≥75 & Driver Gender=M & Pedestrian Gender=M	0.23	9.82	4.84	1.12
223	Pedestrian Age≥75 & Driver Gender=M & Junction Detail=Not at junction	0.19	9.55	4.71	1.09
224	Pedestrian Age≥75 & Driver Gender=M & Vehicle Location=Not at junction	0.19	9.54	4.70	1.09
225	Pedestrian Age≥75 & Driver Gender=M & Junction Control=Not at junction or within 20 metres	0.19	9.47	4.67	1.08
226	Pedestrian Age≥75 & Driver Gender=M & Weather=Fine no high winds	0.38	9.23	4.55	1.06
227	Pedestrian Age≥75 & Pedestrian Gender=M	0.31	8.47	4.17	1.13
228	Pedestrian Age≥75 & Pedestrian Gender=M & Driver IMD Decile=More deprived	0.13	9.47	4.67	1.12
229	Pedestrian Age≥75 & Pedestrian Gender=M & Vehicle Propulsion Code=Petrol	0.14	9.33	4.60	1.10
230	Pedestrian Age≥75 & Pedestrian Gender=M & Driver Home Area=Urban	0.21	9.14	4.51	1.08

ID rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent Antecents	S %	C %	Lift	LIC
231	Pedestrian Age≥75 & Pedestrian Gender=M & Junction Detail=Not at junction	0.14	9.12	4.50	1.08
232	Pedestrian Age≥75 & Pedestrian Gender=M & Vehicle Location=Not at junction	0.13	9.04	4.46	1.07
233	Pedestrian Age≥75 & Pedestrian Gender=M & Vehicle Age≥15	0.23	9.01	4.44	1.06
234	Pedestrian Age≥75 & Pedestrian Gender=M & Weather=Fine no high winds	0.27	9.00	4.44	1.06
235	Pedestrian Age≥75 & Pedestrian Gender=M & Junction Control=Not at junction or within 20 metres	0.14	8.98	4.43	1.06
236	Pedestrian Age≥75 & Pedestrian Gender=M & Driver IMD Decile=Less deprived	0.10	8.90	4.39	1.05
237	Pedestrian Age≥75 & Junction Detail=Not at junction	0.26	8.33	4.11	1.12
238	Pedestrian Age≥75 & Junction Detail=Not at junction & Vehicle Propulsion Code=Petrol	0.11	8.99	4.43	1.08
239	Pedestrian Age≥75 & Junction Detail=Not at junction & Driver Home Area=Urban	0.17	8.98	4.43	1.08
240	Pedestrian Age≥75 & Junction Detail=Not at junction & Vehicle Age≥15	0.19	8.85	4.36	1.06
241	Pedestrian Age≥75 & Vehicle Location=Not at junction	0.26	8.28	4.08	1.11
242	Pedestrian Age≥75 & Vehicle Location=Not at junction & Vehicle Propulsion Code=Petrol	0.11	8.98	4.43	1.08
243	Pedestrian Age≥75 & Vehicle Location=Not at junction & Driver Home Area=Urban	0.17	8.98	4.43	1.08
244	Pedestrian Age≥75 & Vehicle Location=Not at junction & Vehicle Age≥15	0.19	8.85	4.36	1.07
245	Pedestrian Age≥75 & Junction Control=Not at junction or within 20 metres	0.26	8.22	4.05	1.10
246	Pedestrian Age≥75 & Junction Control=Not at junction or within 20 metres & Vehicle Propulsion Code=Petrol	0.11	8.94	4.41	1.09
247	Pedestrian Age≥75 & Junction Control=Not at junction or within 20 metres & Driver Home Area=Urban	0.17	8.92	4.40	1.09
248	Pedestrian Age≥75 & Junction Control=Not at junction or within 20 metres & Vehicle Age≥15	0.19	8.75	4.31	1.06
249	Pedestrian Age≥75 & Junction Control=Not at junction or within 20 metres & Weather=Fine no high winds	0.23	8.66	4.27	1.05
250	Pedestrian Age≥75 & Driver Age=25-34	0.11	8.10	3.99	1.09
251	Pedestrian Age≥75 & Weather=Fine no high winds	0.50	7.95	3.92	1.07
252	Pedestrian Age≥75 & Driver Home Area=Urban	0.38	7.85	3.87	1.05
253	Pedestrian Age≥75 & Driver IMD Decile=Less deprived	0.19	7.84	3.86	1.05
254	Pedestrian Age≥75 & Driver IMD Decile=Less deprived & Road Type=Single carriageway	0.18	8.71	4.30	1.11

Rules with serious injury as consequent

Table 149 – Association rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent, Great Britain..

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
1	Number of Pedestrians involved>2	0.14	42.48	1.75	n.a.
2	Number of Pedestrians involved>2 & Pavement=Dry	0.12	46.20	1.90	1.09
3	Number of Pedestrians involved>2 & Pavement=Dry & Road Type=Single carriageway	0.11	49.31	2.03	1.07
4	Number of Pedestrians involved>2 & Road Type=Single carriageway	0.13	45.45	1.87	1.07
5	Number of Pedestrians involved>2 & Road Type=Single carriageway & Weather=Fine no high winds	0.12	49.06	2.02	1.08
6	Number of Pedestrians involved>2 & Weather=Fine no high winds	0.13	45.13	1.86	1.06
7	Pedestrian Age≥75	2.82	37.35	1.54	n.a.
8	Pedestrian Age≥75 & Vehicle Age≥15	0.18	46.88	1.93	1.26
9	Pedestrian Age≥75 & Vehicle Age≥15 & Pedestrian Gender=F	0.10	52.67	2.17	1.12
10	Pedestrian Age≥75 & Vehicle Age≥15 & Driver Home Area=Urban	0.13	51.14	2.11	1.09
11	Pedestrian Age≥75 & X2nd Road Class=B	0.12	45.35	1.87	1.21
12	Pedestrian Age≥75 & Driver Journey Purpose=Commuting to/from work	0.26	44.53	1.83	1.19
13	Pedestrian Age≥75 & Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead	0.14	49.21	2.03	1.11
14	Pedestrian Age≥75 & Driver Journey Purpose=Commuting to/from work & Driver IMD Decile=More deprived	0.11	46.84	1.93	1.05
15	Pedestrian Age≥75 & Pavement=Wet or damp	0.74	42.93	1.77	1.15
16	Pedestrian Age≥75 & Pavement=Wet or damp & Vehicle Propulsion Code=Petrol	0.37	47.42	1.95	1.10
17	Pedestrian Age≥75 & Pavement=Wet or damp & Driver Gender=F	0.24	46.82	1.93	1.09
18	Pedestrian Age≥75 & Pavement=Wet or damp & Vehicle Manoeuvre=Going ahead	0.35	46.61	1.92	1.09
19	Pedestrian Age≥75 & Pavement=Wet or damp & Pedestrian Gender=F	0.39	45.53	1.87	1.06
20	Pedestrian Age≥75 & Pavement=Wet or damp & Lighting=Darkness - lights lit	0.24	45.51	1.87	1.06
21	Pedestrian Age≥75 & Pavement=Wet or damp & Driver Age=45-54	0.12	45.30	1.87	1.06
22	Pedestrian Age≥75 & Vehicle Engine Capacity ≤1000	0.23	42.66	1.76	1.14
23	Pedestrian Age≥75 & Vehicle Engine Capacity ≤1000 & Vehicle Manoeuvre=Going ahead	0.15	48.53	2.00	1.14
24	Pedestrian Age≥75 & Vehicle Engine Capacity ≤1000 & Driver IMD Decile=More deprived	0.11	48.34	1.99	1.13
25	Pedestrian Age≥75 & Vehicle Engine Capacity ≤1000 & Driver Home Area=Urban	0.17	44.96	1.85	1.05
26	Pedestrian Age≥75 & Vehicle Engine Capacity ≤1000 & Area=Urban	0.21	44.87	1.85	1.05
27	Pedestrian Age≥75 & Driver Age≥75	0.29	42.49	1.75	1.14
28	Pedestrian Age≥75 & Driver Age≥75 & Vehicle Manoeuvre=Going ahead	0.12	47.46	1.95	1.12
29	Pedestrian Age≥75 & Driver Age≥75 & Pedestrian Gender=F	0.20	46.98	1.93	1.11
30	Pedestrian Age≥75 & Driver Home Area=Small town	0.22	42.30	1.74	1.13
31	Pedestrian Age≥75 & Driver Home Area=Small town & Pedestrian Gender=F	0.14	51.40	2.12	1.22
32	Pedestrian Age≥75 & Driver Home Area=Small town & Vehicle 1st Point of Impact=Front	0.11	47.17	1.94	1.12
33	Pedestrian Age≥75 & Driver Home Area=Small town & Area=Rural	0.13	44.56	1.83	1.05

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
34	Pedestrian Age≥75 & Driver Home Area=Small town & Road Type=Single carriageway	0.20	44.48	1.83	1.05
35	Pedestrian Age≥75 & Pedestrian Crossing Physical Facilities=Central refuge	0.18	42.29	1.74	1.13
36	Pedestrian Age≥75 & Weather=Raining no high winds	0.31	42.14	1.73	1.13
37	Pedestrian Age≥75 & Weather=Raining no high winds & Driver Gender=F	0.11	49.66	2.04	1.18
38	Pedestrian Age≥75 & Weather=Raining no high winds & Lighting=Darkness - lights lit	0.12	48.80	2.01	1.16
39	Pedestrian Age≥75 & Weather=Raining no high winds & Vehicle Manoeuvre=Going ahead	0.16	48.40	1.99	1.15
40	Pedestrian Age≥75 & Weather=Raining no high winds & Junction Control=Not at junction or within 20 metres	0.14	47.26	1.95	1.12
41	Pedestrian Age≥75 & Weather=Raining no high winds & Junction Detail=Not at junction	0.14	47.24	1.94	1.12
42	Pedestrian Age≥75 & Weather=Raining no high winds & Vehicle Location=Not at junction	0.14	46.97	1.93	1.11
43	Pedestrian Age≥75 & Weather=Raining no high winds & Vehicle Propulsion Code=Petrol	0.15	46.12	1.90	1.09
44	Pedestrian Age≥75 & Weather=Raining no high winds & Pedestrian Gender=F	0.18	44.87	1.85	1.06
45	Pedestrian Age≥75 & Pedestrian Crossing Physical Facilities=Zebra	0.20	41.77	1.72	1.12
46	Pedestrian Age≥75 & Pedestrian Crossing Physical Facilities=Zebra & Driver Gender=M	0.15	46.70	1.92	1.12
47	Pedestrian Age≥75 & Pedestrian Crossing Physical Facilities=Zebra & Vehicle 1st Point of Impact=Front	0.12	45.65	1.88	1.09
48	Pedestrian Age≥75 & Number of Vehicles=2	0.11	41.04	1.69	1.10
49	Pedestrian Age≥75 & Vehicle Type=Van	0.27	40.77	1.68	1.09
50	Pedestrian Age≥75 & Vehicle Type=Van & Junction Detail=T or staggered junction	0.11	48.10	1.98	1.18
51	Pedestrian Age≥75 & Vehicle Type=Van & Vehicle Location=At junction	0.18	46.18	1.90	1.13
52	Pedestrian Age≥75 & Vehicle Type=Van & Driver IMD Decile=More deprived	0.12	45.45	1.87	1.11
53	Pedestrian Age≥75 & Vehicle Type=Van & Vehicle 1st Point of Impact=Front	0.10	45.39	1.87	1.11
54	Pedestrian Age≥75 & Vehicle Type=Van & Junction Control=Give way/uncontrolled	0.15	45.02	1.85	1.10
55	Pedestrian Age≥75 & Vehicle Propulsion Code=Petrol	1.23	40.68	1.67	1.09
56	Pedestrian Age≥75 & Vehicle Propulsion Code=Petrol & Pedestrian Gender=F	0.69	44.57	1.84	1.10
57	Pedestrian Age≥75 & Vehicle Propulsion Code=Petrol & Driver Age=25-34	0.22	43.28	1.78	1.06
58	Pedestrian Age≥75 & Vehicle Propulsion Code=Petrol & Vehicle Manoeuvre=Going ahead	0.54	43.13	1.78	1.06
59	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead	1.21	40.59	1.67	1.09
60	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Vehicle Engine Capacity =2000-3000	0.13	46.96	1.93	1.16
61	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver Age=25-34	0.24	44.75	1.84	1.10
62	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Pedestrian Gender=F	0.63	43.96	1.81	1.08
63	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver Age=35-44	0.19	43.94	1.81	1.08
64	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & 1st Road Class=B	0.17	43.92	1.81	1.08
65	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver Gender=F	0.33	43.76	1.80	1.08
66	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Driver IMD Decile=More deprived	0.46	43.11	1.78	1.06
67	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing	0.18	43.06	1.77	1.06
68	Pedestrian Age≥75 & Vehicle Manoeuvre=Going ahead & Junction Control=Give way/uncontrolled	0.48	42.90	1.77	1.06
69	Pedestrian Age≥75 & Pedestrian Gender=F	1.58	40.54	1.67	1.09

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
70	Pedestrian Age≥75 & Pedestrian Gender=F & Driver Home Area=Rural	0.16	47.16	1.94	1.16
71	Pedestrian Age≥75 & Pedestrian Gender=F & Junction Detail=Crossroads	0.13	44.97	1.85	1.11
72	Pedestrian Age≥75 & Pedestrian Gender=F & Driver Age=65-74	0.18	44.91	1.85	1.11
73	Pedestrian Age≥75 & Pedestrian Gender=F & Area=Rural	0.24	44.24	1.82	1.09
74	Pedestrian Age≥75 & Pedestrian Gender=F & Driver Age=0-24	0.14	43.96	1.81	1.08
75	Pedestrian Age≥75 & Pedestrian Gender=F & Driver IMD Decile=Less deprived	0.56	43.71	1.80	1.08
76	Pedestrian Age≥75 & Pedestrian Gender=F & Vehicle Engine Capacity =1000-1500	0.37	43.53	1.79	1.07
77	Pedestrian Age≥75 & Pedestrian Gender=F & Vehicle 1st Point of Impact=Nearside/Offside	0.35	43.17	1.78	1.06
78	Pedestrian Age≥75 & Pedestrian Gender=F & Vehicle Engine Capacity =1500-2000	0.56	43.15	1.78	1.06
79	Pedestrian Age≥75 & Pedestrian Gender=F & Driver Gender=F	0.48	42.99	1.77	1.06
80	Pedestrian Age≥75 & Driver Home Area=Rural	0.25	40.29	1.66	1.08
81	Pedestrian Age≥75 & Driver Home Area=Rural & Driver IMD Decile=Less deprived	0.15	43.04	1.77	1.07
82	Pedestrian Age≥75 & Driver Gender=F	0.82	39.86	1.64	1.07
83	Pedestrian Age≥75 & Driver Gender=F & Area=Rural	0.15	44.84	1.85	1.13
84	Pedestrian Age≥75 & Driver Gender=F & Lighting=Darkness - lights lit	0.12	44.13	1.82	1.11
85	Pedestrian Age≥75 & Driver Gender=F & Vehicle 1st Point of Impact=Nearside/Offside	0.21	43.99	1.81	1.10
86	Pedestrian Age≥75 & Driver Gender=F & Driver Age=55-64	0.11	42.53	1.75	1.07
87	Pedestrian Age≥75 & Vehicle Engine Capacity =1000-1500	0.66	39.79	1.64	1.07
88	Pedestrian Age≥75 & Vehicle Engine Capacity =1000-1500 & Vehicle 1st Point of Impact=Nearside/Offside	0.18	44.36	1.83	1.12
89	Pedestrian Age≥75 & Vehicle Engine Capacity =1000-1500 & Area=Rural	0.12	44.32	1.82	1.11
90	Pedestrian Age≥75 & Vehicle Engine Capacity =1000-1500 & Vehicle Manoeuvre=Reversing	0.14	43.30	1.78	1.09
91	Pedestrian Age≥75 & Vehicle Engine Capacity =1000-1500 & Driver Age=25-34	0.10	41.92	1.73	1.05
92	Pedestrian Age≥75 & Vehicle Engine Capacity =1000-1500 & Vehicle 1st Point of Impact=Back	0.13	41.78	1.72	1.05
93	Pedestrian Age≥75 & Driver Age=55-64	0.37	39.68	1.63	1.06
94	Pedestrian Age≥75 & Driver Age=55-64 & Driver Journey Purpose=Journey as part of work	0.13	43.14	1.78	1.09
95	Pedestrian Age≥75 & Driver Age=55-64 & Vehicle Engine Capacity =1500-2000	0.15	42.86	1.76	1.08
96	Pedestrian Age≥75 & Junction Detail=Crossroads	0.23	39.49	1.63	1.06
97	Pedestrian Age≥75 & Junction Detail=Crossroads & Road Type=Single carriageway	0.19	41.64	1.71	1.05
98	Pedestrian Age≥75 & Vehicle Engine Capacity =2000-3000	0.32	39.44	1.62	1.06
99	Pedestrian Age≥75 & Vehicle Engine Capacity =2000-3000 & 1st Road Class=A	0.12	50.97	2.10	1.29
100	Pedestrian Age≥75 & Vehicle Engine Capacity =2000-3000 & Driver IMD Decile=Less deprived	0.13	43.35	1.78	1.10
101	Pedestrian Age≥75 & Vehicle Engine Capacity =2000-3000 & Driver Home Area=Urban	0.23	42.70	1.76	1.08
102	Pedestrian Age≥75 & Vehicle Engine Capacity =2000-3000 & Driver IMD Decile=More deprived	0.14	42.40	1.75	1.07
103	Pedestrian Age≥75 & Driver IMD Decile=Less deprived	0.96	39.44	1.62	1.06
104	Pedestrian Age≥75 & Driver IMD Decile=Less deprived & Driver Age=25-34	0.16	44.03	1.81	1.12
105	Pedestrian Age≥75 & Driver IMD Decile=Less deprived & Driver Age=65-74	0.14	42.01	1.73	1.07
106	Pedestrian Age≥75 & Driver IMD Decile=Less deprived & Vehicle 1st Point of Impact=Front	0.49	41.79	1.72	1.06

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
107	Pedestrian Age≥75 & Lighting=Darkness - lights lit	0.43	39.43	1.62	1.06
108	Pedestrian Age≥75 & Lighting=Darkness - lights lit & Driver IMD Decile=More deprived	0.19	44.41	1.83	1.13
109	Pedestrian Age≥75 & Driver IMD Decile=More deprived	1.11	39.38	1.62	1.05
110	Pedestrian Age≥75 & Driver IMD Decile=More deprived & Area=Rural	0.13	44.33	1.83	1.13
111	Pedestrian Age≥75 & Driver IMD Decile=More deprived & Vehicle 1st Point of Impact=Nearside/Offside	0.24	42.16	1.74	1.07
112	Pedestrian Age≥75 & Driver IMD Decile=More deprived & Driver Journey Purpose=Journey as part of work	0.30	41.49	1.71	1.05
113	Pedestrian Age=65-74	2.22	33.41	1.38	n.a.
114	Pedestrian Age=65-74 & Driver Journey Purpose=Commuting to/from work	0.21	42.22	1.74	1.26
115	Pedestrian Age=65-74 & Driver Journey Purpose=Commuting to/from work & Pedestrian Gender=F	0.11	46.95	1.93	1.11
116	Pedestrian Age=65-74 & Driver Journey Purpose=Commuting to/from work & Driver IMD Decile=More deprived	0.10	46.31	1.91	1.10
117	Pedestrian Age=65-74 & Driver Journey Purpose=Commuting to/from work & Driver Home Area=Urban	0.18	45.04	1.85	1.07
118	Pedestrian Age=65-74 & Driver Journey Purpose=Commuting to/from work & Vehicle 1st Point of Impact=Front	0.16	44.68	1.84	1.06
119	Pedestrian Age=65-74 & Driver Age=0-24	0.27	39.57	1.63	1.18
120	Pedestrian Age=65-74 & Driver Age=0-24 & Vehicle Propulsion Code=Petrol	0.17	45.49	1.87	1.15
121	Pedestrian Age=65-74 & Driver Age=0-24 & Vehicle Manoeuvre=Going ahead	0.17	44.40	1.83	1.12
122	Pedestrian Age=65-74 & Driver Age=0-24 & Pedestrian Gender=F	0.15	43.91	1.81	1.11
123	Pedestrian Age=65-74 & Driver Age=0-24 & Driver IMD Decile=More deprived	0.13	43.27	1.78	1.09
124	Pedestrian Age=65-74 & Driver Age=0-24 & Vehicle Age≥15	0.22	42.44	1.75	1.07
125	Pedestrian Age=65-74 & Driver Age=0-24 & Vehicle 1st Point of Impact=Front	0.18	42.27	1.74	1.07
126	Pedestrian Age=65-74 & Lighting=Darkness - lights lit	0.54	38.75	1.60	1.16
127	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Driver Gender=F	0.12	46.82	1.93	1.21
128	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Vehicle Propulsion Code=Petrol	0.26	44.59	1.84	1.15
129	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Vehicle Engine Capacity =1000-1500	0.12	44.51	1.83	1.15
130	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Vehicle Manoeuvre=Going ahead	0.30	43.01	1.77	1.11
131	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Driver Home Area=Urban	0.40	42.81	1.76	1.10
132	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Driver IMD Decile=More deprived	0.24	42.56	1.75	1.10
133	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Pedestrian Gender=F	0.24	41.86	1.72	1.08
134	Pedestrian Age=65-74 & Lighting=Darkness - lights lit & Vehicle 1st Point of Impact=Nearside/Offside	0.14	41.41	1.70	1.07
135	Pedestrian Age=65-74 & Vehicle Age≥15	0.13	38.39	1.58	1.15
136	Pedestrian Age=65-74 & Vehicle Age≥15 & Driver Gender=M	0.11	41.04	1.69	1.07
137	Pedestrian Age=65-74 & Vehicle Age≥15 & Area=Urban	0.12	40.93	1.69	1.07
138	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500	0.49	37.68	1.55	1.13
139	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Day of Week=Weekend	0.13	44.55	1.83	1.18
140	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Pavement=Wet or damp	0.16	42.80	1.76	1.14
141	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Driver IMD Decile=More deprived	0.23	42.05	1.73	1.12
142	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Vehicle Manoeuvre=Going ahead	0.23	41.94	1.73	1.11
143	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & 1st Road Class=A	0.15	41.53	1.71	1.10

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
144	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Vehicle 1st Point of Impact=Front	0.31	41.53	1.71	1.10
145	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Pedestrian Gender=F	0.27	41.11	1.69	1.09
146	Pedestrian Age=65-74 & Vehicle Engine Capacity =1000-1500 & Driver Home Area=Urban	0.36	39.57	1.63	1.05
147	Pedestrian Age=65-74 & Pavement=Wet or damp	0.63	37.63	1.55	1.13
148	Pedestrian Age=65-74 & Pavement=Wet or damp & Vehicle Manoeuvre=Going ahead	0.33	43.52	1.79	1.16
149	Pedestrian Age=65-74 & Pavement=Wet or damp & Driver Gender=F	0.18	41.87	1.72	1.11
150	Pedestrian Age=65-74 & Pavement=Wet or damp & Pedestrian Gender=F	0.33	41.37	1.70	1.10
151	Pedestrian Age=65-74 & Pavement=Wet or damp & Vehicle Propulsion Code=Petrol	0.29	41.34	1.70	1.10
152	Pedestrian Age=65-74 & Pavement=Wet or damp & Weather=Fine no high winds	0.30	40.49	1.67	1.08
153	Pedestrian Age=65-74 & Pavement=Wet or damp & Driver Age=45-54	0.12	40.10	1.65	1.07
154	Pedestrian Age=65-74 & Pavement=Wet or damp & Vehicle Location=Not at junction	0.23	40.05	1.65	1.06
155	Pedestrian Age=65-74 & Pavement=Wet or damp & Junction Detail=Not at junction	0.23	40.05	1.65	1.06
156	Pedestrian Age=65-74 & Pavement=Wet or damp & Junction Control=Not at junction or within 20 metres	0.24	39.75	1.64	1.06
157	Pedestrian Age=65-74 & Pavement=Wet or damp & Vehicle 1st Point of Impact=Front	0.42	39.69	1.63	1.05
158	Pedestrian Age=65-74 & Driver Home Area=Rural	0.18	37.54	1.55	1.12
159	Pedestrian Age=65-74 & Road Type=Dual carriageway	0.18	36.89	1.52	1.10
160	Pedestrian Age=65-74 & Road Type=Dual carriageway & Vehicle Manoeuvre=Going ahead	0.11	41.85	1.72	1.13
161	Pedestrian Age=65-74 & Road Type=Dual carriageway & Driver Gender=M	0.13	39.47	1.63	1.07
162	Pedestrian Age=65-74 & Road Type=Dual carriageway & Driver Home Area=Urban	0.12	38.97	1.60	1.06
163	Pedestrian Age=65-74 & Driver Age=35-44	0.36	36.35	1.50	1.09
164	Pedestrian Age=65-74 & Driver Age=35-44 & Vehicle Manoeuvre=Going ahead	0.17	40.79	1.68	1.12
165	Pedestrian Age=65-74 & Driver Age=35-44 & Vehicle Manoeuvre=Turning left/right/U	0.11	38.80	1.60	1.07
166	Pedestrian Age=65-74 & Driver Age=35-44 & Vehicle Propulsion Code=Petrol	0.14	38.59	1.59	1.06
167	Pedestrian Age=65-74 & Driver Age=35-44 & Pedestrian Gender=F	0.19	38.46	1.58	1.06
168	Pedestrian Age=65-74 & Driver Age=35-44 & Vehicle 1st Point of Impact=Front	0.21	38.38	1.58	1.06
169	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol	0.93	36.24	1.49	1.08
170	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Vehicle Manoeuvre=Going ahead	0.47	41.04	1.69	1.13
171	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Day of Week=Weekend	0.23	39.64	1.63	1.09
172	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Driver IMD Decile=More deprived	0.42	38.97	1.60	1.08
173	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Vehicle 1st Point of Impact=Front	0.58	38.92	1.60	1.07
174	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Driver Age=25-34	0.18	38.73	1.59	1.07
175	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Pedestrian Gender=F	0.48	38.60	1.59	1.07
176	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & 1st Road Class=A	0.30	38.39	1.58	1.06
177	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing	0.11	38.22	1.57	1.05
178	Pedestrian Age=65-74 & Vehicle Propulsion Code=Petrol & Driver Home Area=Urban	0.69	38.06	1.57	1.05
179	Pedestrian Age=65-74 & Driver Age=25-34	0.42	36.22	1.49	1.08

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
180	Pedestrian Age=65-74 & Driver Age=25-34 & Vehicle Manoeuvre=Going ahead	0.21	40.23	1.66	1.11
181	Pedestrian Age=65-74 & Driver Age=25-34 & Driver IMD Decile=Less deprived	0.11	39.46	1.62	1.09
182	Pedestrian Age=65-74 & Driver Age=25-34 & Vehicle 1st Point of Impact=Front	0.27	39.15	1.61	1.08
183	Pedestrian Age=65-74 & Number of Vehicles=2	0.13	36.21	1.49	1.08
184	Pedestrian Age=65-74 & Number of Vehicles=2 & Pavement=Dry	0.10	38.64	1.59	1.07
185	Pedestrian Age=65-74 & Pedestrian Crossing Physical Facilities=Pedestrian phase at traffic signal junction	0.24	36.08	1.49	1.08
186	Pedestrian Age=65-74 & Pedestrian Crossing Physical Facilities=Pedestrian phase at traffic signal junction & Vehicle Manoeuvre=Going ahead	0.12	39.90	1.64	1.11
187	Pedestrian Age=65-74 & Pedestrian Crossing Physical Facilities=Pedestrian phase at traffic signal junction & Vehicle Age≥15	0.17	39.26	1.62	1.09
188	Pedestrian Age=65-74 & Pedestrian Crossing Physical Facilities=Pedestrian phase at traffic signal junction & Driver IMD Decile=More deprived	0.12	39.22	1.61	1.09
189	Pedestrian Age=65-74 & Pedestrian Crossing Physical Facilities=Pedestrian phase at traffic signal junction & Vehicle 1st Point of Impact=Front	0.16	39.08	1.61	1.08
190	Pedestrian Age=65-74 & Driver Home Area=Urban	1.48	35.79	1.47	1.07
191	Pedestrian Age=65-74 & Driver Home Area=Urban & Vehicle Manoeuvre=Going ahead	0.69	39.68	1.63	1.11
192	Pedestrian Age=65-74 & Driver Home Area=Urban & Day of Week=Weekend	0.34	39.48	1.63	1.10
193	Pedestrian Age=65-74 & Driver Home Area=Urban & Vehicle Engine Capacity ≤1000	0.12	39.05	1.61	1.09
194	Pedestrian Age=65-74 & Driver Home Area=Urban & Vehicle 1st Point of Impact=Front	0.92	38.76	1.60	1.08
195	Pedestrian Age=65-74 & Driver Home Area=Urban & Pedestrian Gender=F	0.77	37.84	1.56	1.06
196	Pedestrian Age=65-74 & Driver IMD Decile=More deprived	0.92	35.74	1.47	1.07
197	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Vehicle Manoeuvre=Turning left/right/U	0.26	39.73	1.64	1.11
198	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Vehicle Manoeuvre=Going ahead	0.42	38.61	1.59	1.08
199	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Junction Detail=Crossroads	0.11	38.58	1.59	1.08
200	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Junction Control=Auto traffic signal	0.13	38.53	1.59	1.08
201	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Day of Week=Weekend	0.21	37.98	1.56	1.06
202	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Vehicle 1st Point of Impact=Front	0.56	37.88	1.56	1.06
203	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Vehicle Engine Capacity =2000-3000	0.11	37.69	1.55	1.05
204	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Junction Detail=T or staggered junction	0.38	37.68	1.55	1.05
205	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & 1st Road Class=B	0.10	37.57	1.55	1.05
206	Pedestrian Age=65-74 & Driver IMD Decile=More deprived & Vehicle Location=At junction	0.62	37.55	1.55	1.05
207	Pedestrian Age=65-74 & Vehicle Manoeuvre=Turning left/right/U	0.52	35.69	1.47	1.07
208	Pedestrian Age=65-74 & Vehicle Manoeuvre=Turning left/right/U & Day of Week=Weekend	0.11	39.13	1.61	1.10
209	Pedestrian Age=65-74 & Vehicle Manoeuvre=Turning left/right/U & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.37	37.71	1.55	1.06
210	Pedestrian Age=65-74 & Vehicle 1st Point of Impact=Front	1.30	35.67	1.47	1.07
211	Pedestrian Age=65-74 & Vehicle 1st Point of Impact=Front & Vehicle Engine Capacity ≤1000	0.12	41.75	1.72	1.17
212	Pedestrian Age=65-74 & Vehicle 1st Point of Impact=Front & Driver Age=55-64	0.18	41.24	1.70	1.16
213	Pedestrian Age=65-74 & Vehicle 1st Point of Impact=Front & Pedestrian Gender=F	0.68	38.09	1.57	1.07
214	Pedestrian Age=65-74 & Vehicle 1st Point of Impact=Front & Day of Week=Weekend	0.31	37.91	1.56	1.06
215	Pedestrian Age=65-74 & Vehicle 1st Point of Impact=Front & Vehicle Manoeuvre=Going ahead	0.74	37.64	1.55	1.06
216	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead	1.06	35.60	1.47	1.07

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
217	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Weather=Raining no high winds	0.14	41.41	1.70	1.16
218	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Driver Journey Purpose=Journey as part of work	0.19	39.51	1.63	1.11
219	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing	0.16	38.87	1.60	1.09
220	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Driver Age=55-64	0.12	38.76	1.60	1.09
221	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Driver Gender=F	0.24	38.70	1.59	1.09
222	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Day of Week=Weekend	0.26	38.44	1.58	1.08
223	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Pedestrian Gender=F	0.52	38.22	1.57	1.07
224	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Vehicle Age≥15	0.72	38.00	1.56	1.07
225	Pedestrian Age=65-74 & Vehicle Manoeuvre=Going ahead & Driver IMD Decile=Less deprived	0.31	37.73	1.55	1.06
226	Pedestrian Age=65-74 & Pedestrian Gender=F	1.16	35.55	1.46	1.06
227	Pedestrian Age=65-74 & Pedestrian Gender=F & Junction Detail=Crossroads	0.11	41.44	1.71	1.17
228	Pedestrian Age=65-74 & Pedestrian Gender=F & Driver Age=55-64	0.15	41.35	1.70	1.16
229	Pedestrian Age=65-74 & Pedestrian Gender=F & Driver IMD Decile=Less deprived	0.38	39.20	1.61	1.10
230	Pedestrian Age=65-74 & Pedestrian Gender=F & Weather=Raining no high winds	0.15	39.04	1.61	1.10
231	Pedestrian Age=65-74 & Pedestrian Gender=F & Vehicle Engine Capacity =2000-3000	0.13	38.99	1.61	1.10
232	Pedestrian Age=65-74 & Pedestrian Gender=F & 1st Road Class=B	0.15	37.55	1.55	1.06
233	Pedestrian Age=65-74 & Pedestrian Gender=F & Driver Gender=M	0.78	37.37	1.54	1.05
234	Pedestrian Age=65-74 & Day of Week=Weekend	0.49	35.51	1.46	1.06
235	Pedestrian Age=65-74 & Day of Week=Weekend & Driver IMD Decile=Less deprived	0.16	40.86	1.68	1.15
236	Pedestrian Age=65-74 & Day of Week=Weekend & Junction Control=Give way/uncontrolled	0.24	38.02	1.57	1.07
237	Pedestrian Age=65-74 & Day of Week=Weekend & Driver Gender=F	0.11	37.89	1.56	1.07
238	Pedestrian Age=65-74 & Day of Week=Weekend & Vehicle 1st Point of Impact=Nearside/Offside	0.12	37.50	1.54	1.06
239	Pedestrian Age=65-74 & Driver Age=55-64	0.27	35.49	1.46	1.06
240	Pedestrian Age=65-74 & Driver Age=55-64 & Driver IMD Decile=Less deprived	0.11	38.42	1.58	1.08
241	Pedestrian Age=65-74 & Driver Age=55-64 & Vehicle Location=Not at junction	0.10	37.70	1.55	1.06
242	Pedestrian Age=65-74 & Driver Age=55-64 & Junction Detail=Not at junction	0.10	37.70	1.55	1.06
243	Pedestrian Age=65-74 & Driver Age=55-64 & Junction Control=Not at junction or within 20 metres	0.10	37.50	1.54	1.06
244	Pedestrian Age=65-74 & 1st Road Class=B	0.27	35.35	1.46	1.06
245	Pedestrian Age=65-74 & 1st Road Class=B & Vehicle Engine Capacity =1500-2000	0.11	40.34	1.66	1.14
246	Number of Pedestrians involved=2	0.85	30.79	1.27	n.a.
247	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived	0.26	36.12	1.49	1.17
248	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived & Vehicle Location=Not at junction	0.12	40.31	1.66	1.12
249	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived & Junction Detail=Not at junction	0.12	40.31	1.66	1.12
250	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived & Junction Control=Not at junction or within 20 metres	0.12	39.50	1.63	1.09
251	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.18	39.20	1.61	1.09
252	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived & Vehicle Manoeuvre=Going ahead	0.15	38.67	1.59	1.07

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
253	Number of Pedestrians involved=2 & Driver IMD Decile=Less deprived & Weather=Fine no high winds	0.23	38.26	1.58	1.06
254	Number of Pedestrians involved=2 & Area=Rural	0.11	35.98	1.48	1.17
255	Number of Pedestrians involved=2 & Area=Rural & Weather=Fine no high winds	0.10	39.77	1.64	1.11
256	Number of Pedestrians involved=2 & 1st Road Class=B	0.10	35.71	1.47	1.16
257	Number of Pedestrians involved=2 & 1st Road Class=B & Road Type=Single carriageway	0.10	39.43	1.62	1.10
258	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead	0.49	35.35	1.46	1.15
259	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead & Vehicle Engine Capacity =1500-2000	0.20	40.06	1.65	1.13
260	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead & Day of Week=Weekend	0.15	39.61	1.63	1.12
261	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead & Vehicle Propulsion Code=Heavy oil	0.17	38.67	1.59	1.09
262	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead & 1st Road Class=A	0.21	37.99	1.56	1.07
263	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead & Vehicle 1st Point of Impact=Front	0.39	37.61	1.55	1.06
264	Number of Pedestrians involved=2 & Vehicle Manoeuvre=Going ahead & Driver Gender=M	0.33	37.38	1.54	1.06
265	Number of Pedestrians involved=2 & Day of Week=Weekend	0.25	34.73	1.43	1.13
266	Number of Pedestrians involved=2 & Day of Week=Weekend & 1st Road Class=A	0.11	40.54	1.67	1.17
267	Number of Pedestrians involved=2 & Day of Week=Weekend & Driver Home Area=Urban	0.16	38.41	1.58	1.11
268	Number of Pedestrians involved=2 & Day of Week=Weekend & Junction Control=Not at junction or within 20 metres	0.12	37.22	1.53	1.07
269	Number of Pedestrians involved=2 & Day of Week=Weekend & Junction Detail=Not at junction	0.12	36.99	1.52	1.07
270	Number of Pedestrians involved=2 & Day of Week=Weekend & Vehicle Location=Not at junction	0.12	36.82	1.52	1.06
271	Number of Pedestrians involved=2 & Driver Age=45-54	0.14	34.70	1.43	1.13
272	Number of Pedestrians involved=2 & Driver Age=45-54 & Pavement=Dry	0.11	40.21	1.66	1.16
273	Number of Pedestrians involved=2 & Driver Age=45-54 & Weather=Fine no high winds	0.12	38.43	1.58	1.11
274	Number of Pedestrians involved=2 & Junction Control=Auto traffic signal	0.10	34.31	1.41	1.11
275	Number of Pedestrians involved=2 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing	0.12	33.61	1.38	1.09
276	Number of Pedestrians involved=2 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing & Weather=Fine no high winds	0.10	36.84	1.52	1.10
277	Number of Pedestrians involved=2 & 1st Road Class=A	0.32	33.38	1.37	1.08
278	Number of Pedestrians involved=2 & 1st Road Class=A & Weather=Fine no high winds	0.29	36.01	1.48	1.08
279	Number of Pedestrians involved=2 & 1st Road Class=A & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.16	35.96	1.48	1.08
280	Number of Pedestrians involved=2 & Driver Gender=F	0.21	33.26	1.37	1.08
281	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000	0.34	33.09	1.36	1.07
282	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Junction Control=Not at junction or within 20 metres	0.15	36.56	1.51	1.10
283	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Lighting=Darkness - lights lit	0.11	36.55	1.50	1.10
284	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Junction Detail=Not at junction	0.15	36.46	1.50	1.10
285	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Vehicle Location=Not at junction	0.15	36.33	1.50	1.10
286	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Vehicle 1st Point of Impact=Front	0.25	36.32	1.50	1.10
287	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Pavement=Dry	0.25	35.11	1.45	1.06
288	Number of Pedestrians involved=2 & Vehicle Engine Capacity =1500-2000 & Weather=Fine no high winds	0.29	34.91	1.44	1.05

ID rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
289	Number of Pedestrians involved=2 & Driver Home Area=Urban	0.56	32.64	1.34	1.06
290	Number of Pedestrians involved=2 & Driver Home Area=Urban & Vehicle Location=Not at junction	0.24	36.36	1.50	1.11
291	Number of Pedestrians involved=2 & Driver Home Area=Urban & Junction Detail=Not at junction	0.24	36.36	1.50	1.11
292	Number of Pedestrians involved=2 & Driver Home Area=Urban & Junction Control=Not at junction or within 20 metres	0.24	36.30	1.49	1.11
293	Number of Pedestrians involved=2 & Driver Home Area=Urban & Weather=Fine no high winds	0.49	34.66	1.43	1.06
294	Number of Pedestrians involved=2 & Junction Detail=Not at junction	0.37	32.59	1.34	1.06
295	Number of Pedestrians involved=2 & Junction Detail=Not at junction & Driver IMD Decile=More deprived	0.16	36.81	1.52	1.13
296	Number of Pedestrians involved=2 & Junction Detail=Not at junction & Vehicle 1st Point of Impact=Front	0.25	36.15	1.49	1.11
297	Number of Pedestrians involved=2 & Junction Detail=Not at junction & Driver Gender=M	0.25	35.26	1.45	1.08
298	Number of Pedestrians involved=2 & Junction Detail=Not at junction & Vehicle Propulsion Code=Petrol	0.16	34.82	1.43	1.07
299	Number of Pedestrians involved=2 & Junction Detail=Not at junction & Weather=Fine no high winds	0.32	34.82	1.43	1.07
300	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres	0.38	32.57	1.34	1.06
301	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres & Driver IMD Decile=More deprived	0.16	37.20	1.53	1.14
302	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres & Vehicle 1st Point of Impact=Front	0.25	35.90	1.48	1.10
303	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres & Driver Gender=M	0.26	35.02	1.44	1.08
304	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres & Vehicle Propulsion Code=Petrol	0.16	34.81	1.43	1.07
305	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres & Weather=Fine no high winds	0.33	34.68	1.43	1.06
306	Number of Pedestrians involved=2 & Junction Control=Not at junction or within 20 metres & Lighting=Darkness - lights lit	0.10	34.52	1.42	1.06
307	Number of Pedestrians involved=2 & Vehicle Location=Not at junction	0.37	32.55	1.34	1.06
308	Number of Pedestrians involved=2 & Vehicle Location=Not at junction & Driver IMD Decile=More deprived	0.16	36.81	1.52	1.13
309	Number of Pedestrians involved=2 & Vehicle Location=Not at junction & Vehicle 1st Point of Impact=Front	0.25	36.15	1.49	1.11
310	Number of Pedestrians involved=2 & Vehicle Location=Not at junction & Driver Gender=M	0.25	35.26	1.45	1.08
311	Number of Pedestrians involved=2 & Vehicle Location=Not at junction & Vehicle Propulsion Code=Petrol	0.16	34.82	1.43	1.07
312	Number of Pedestrians involved=2 & Vehicle Location=Not at junction & Weather=Fine no high winds	0.32	34.77	1.43	1.07

Table 150 – Association rules with vehicle characteristics as first antecedent and serious injury crashes as consequent, Great Britain..

ID rule	Rules with vehicle characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
313	Vehicle Skidding and Overturning=Yes	0.97	35.37	1.46	n.a.
314	Vehicle Skidding and Overturning=Yes & Speed limit≥50	0.12	43.41	1.79	1.23
315	Vehicle Skidding and Overturning=Yes & Speed limit≥50 & Vehicle Towing and Articulation=No tow/articulation	0.11	47.17	1.94	1.09
316	Vehicle Skidding and Overturning=Yes & Road Type=Dual carriageway	0.16	41.83	1.72	1.18
317	Vehicle Skidding and Overturning=Yes & 1st Road Class=B	0.12	40.91	1.68	1.16
318	Vehicle Skidding and Overturning=Yes & 1st Road Class=B & Vehicle Manoeuvre=Going ahead	0.10	45.03	1.85	1.10
319	Vehicle Skidding and Overturning=Yes & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing	0.14	37.85	1.56	1.07
320	Vehicle Skidding and Overturning=Yes & Area=Rural	0.19	37.18	1.53	1.05
321	Vehicle Skidding and Overturning=Yes & Area=Rural & Vehicle Location=Not at junction	0.14	40.08	1.65	1.08
322	Vehicle Skidding and Overturning=Yes & Area=Rural & Junction Detail=Not at junction	0.14	40.08	1.65	1.08
323	Vehicle Skidding and Overturning=Yes & Area=Rural & Junction Control=Not at junction or within 20 metres	0.14	40.08	1.65	1.08
324	Vehicle Skidding and Overturning=Yes & Area=Rural & Vehicle Type=Car	0.14	39.18	1.61	1.05
325	Number of Vehicles≥2	0.34	34.02	1.40	n.a.
326	Vehicle Towing and Articulation=Articulated vehicle	0.16	32.74	1.35	n.a.
327	Vehicle Towing and Articulation=Articulated vehicle & Area=Urban	0.12	36.70	1.51	1.12
328	Vehicle Towing and Articulation=Articulated vehicle & 1st Road Class=A	0.11	35.82	1.47	1.09
329	Vehicle Type=PTW<500	0.31	30.16	1.24	n.a.

Table 151 – Association rules with roadway characteristics as first antecedent and serious injury crashes as consequent, Great Britain..

ID rule	Rules with roadway characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
330	Speed limit=40	1.23	34.73	1.43	n.a.
331	Speed limit=40 & Day of Week=Weekend	0.32	39.63	1.63	1.14
332	Speed limit=40 & Day of Week=Weekend & 1st Road Class=A	0.24	41.88	1.72	1.06
333	Speed limit≥50	1.01	33.33	1.37	n.a.
334	Speed limit≥50 & 1st Road Class=A	0.59	36.72	1.51	1.10
335	Speed limit≥50 & Road Type=Dual carriageway	0.30	36.16	1.49	1.08
336	Speed limit≥50 & Number of Vehicles=2	0.13	35.16	1.45	1.05

Table 152 – Association rules with environmental characteristics as first antecedent and serious injury crashes as consequent, Great Britain..

ID rule	Rules with environmental characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
337	Lighting=Darkness - no lighting	0.61	33.20	1.37	n.a.
338	Lighting=Darkness - no lighting & Junction Detail=T or staggered junction	0.10	37.70	1.55	1.14
339	Lighting=Darkness - no lighting & Speed limit≥50	0.34	35.51	1.46	1.07
340	Lighting=Darkness - no lighting & 1st Road Class=A	0.28	35.22	1.45	1.06
341	Lighting=Darkness - no lighting & Junction Control=Give way/uncontrolled	0.15	35.09	1.44	1.06
342	Weather=Raining + high winds	0.31	31.09	1.28	n.a.
343	Weather=Raining + high winds & Lighting=Darkness - lights lit	0.19	33.69	1.39	1.08
344	Weather=Raining + high winds & Lighting=Darkness - lights lit & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.12	36.73	1.51	1.09
345	Weather=Raining + high winds & Junction Detail=T or staggered junction	0.12	32.94	1.36	1.06
346	Weather=Raining + high winds & Junction Detail=T or staggered junction & Road Type=Single carriageway	0.11	35.35	1.46	1.07
347	Lighting=Darkness - lights unlit	0.22	29.32	1.21	n.a.
348	Lighting=Darkness - lights unlit & Junction Detail=Not at junction	0.13	37.24	1.53	1.27
349	Lighting=Darkness - lights unlit & Junction Control=Not at junction or within 20 metres	0.13	36.18	1.49	1.23

Table 153 – Association rules with driver characteristics as first antecedent and serious injury crashes as consequent, Great Britain..

ID rule	Rules with driver characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
350	Driver Journey Purpose=Commuting to/from work	2.61	30.07	1.24	n.a.
351	Driver Journey Purpose=Commuting to/from work & Speed limit=40	0.20	41.16	1.69	1.37
352	Driver Journey Purpose=Commuting to/from work & Speed limit=40 & Vehicle Manoeuvre=Going ahead	0.17	47.46	1.95	1.15
353	Driver Journey Purpose=Commuting to/from work & Speed limit=40 & Vehicle 1st Point of Impact=Front	0.15	44.44	1.83	1.08
354	Driver Journey Purpose=Commuting to/from work & Speed limit≥50	0.16	39.93	1.64	1.33
355	Driver Journey Purpose=Commuting to/from work & Speed limit≥50 & Road Type=Single carriageway	0.11	44.05	1.81	1.10
356	Driver Journey Purpose=Commuting to/from work & Speed limit≥50 & Area=Rural	0.14	42.27	1.74	1.06
357	Driver Journey Purpose=Commuting to/from work & Speed limit≥50 & Day of Week=Weekday	0.13	41.98	1.73	1.05
358	Driver Journey Purpose=Commuting to/from work & Vehicle Type=Van	0.20	37.95	1.56	1.26
359	Driver Journey Purpose=Commuting to/from work & Vehicle Type=Van & 1st Road Class=A	0.10	44.74	1.84	1.18
360	Driver Journey Purpose=Commuting to/from work & Vehicle Type=Van & Vehicle Manoeuvre=Going ahead	0.13	41.31	1.70	1.09
361	Driver Journey Purpose=Commuting to/from work & Vehicle Type=Van & Weather=Fine no high winds	0.17	40.14	1.65	1.06
362	Driver Journey Purpose=Commuting to/from work & Area=Rural	0.41	36.45	1.50	1.21
363	Driver Journey Purpose=Commuting to/from work & Area=Rural & Pavement=Wet or damp	0.15	40.32	1.66	1.11
364	Driver Journey Purpose=Commuting to/from work & Area=Rural & Vehicle Location=Not at junction	0.27	40.27	1.66	1.10
365	Driver Journey Purpose=Commuting to/from work & Area=Rural & Junction Detail=Not at junction	0.27	40.27	1.66	1.10
366	Driver Journey Purpose=Commuting to/from work & Area=Rural & Junction Control=Not at junction or within 20 metres	0.27	40.09	1.65	1.10

ID rule	Rules with driver characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
367	Driver Journey Purpose=Commuting to/from work & Area=Rural & 1st Road Class=A	0.19	39.50	1.63	1.08
368	Driver Journey Purpose=Commuting to/from work & Area=Rural & Vehicle 1st Point of Impact=Front	0.28	39.08	1.61	1.07
369	Driver Journey Purpose=Commuting to/from work & Area=Rural & Vehicle Manoeuvre=Going ahead	0.31	38.42	1.58	1.05
370	Driver Journey Purpose=Commuting to/from work & Number of Vehicles=2	0.15	34.74	1.43	1.16
371	Driver Journey Purpose=Commuting to/from work & Number of Vehicles=2 & Vehicle 1st Point of Impact=Front	0.10	38.89	1.60	1.12
372	Driver Journey Purpose=Commuting to/from work & Number of Vehicles=2 & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.11	37.69	1.55	1.08
373	Driver Journey Purpose=Commuting to/from work & Road Type=Dual carriageway	0.37	34.50	1.42	1.15
374	Driver Journey Purpose=Commuting to/from work & Road Type=Dual carriageway & Vehicle Manoeuvre=Going ahead	0.27	39.39	1.62	1.14
375	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead	1.82	33.97	1.40	1.13
376	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead & Pedestrian Crossing Physical Facilities=Central refuge	0.11	39.58	1.63	1.17
377	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead & Lighting=Darkness - lights lit	0.53	36.72	1.51	1.08
378	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead & Pavement=Wet or damp	0.56	36.03	1.48	1.06
379	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead & Day of Week=Weekend	0.21	35.95	1.48	1.06
380	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead & Junction Detail=T or staggered junction	0.57	35.79	1.47	1.05
381	Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead & 1st Road Class=A	0.82	35.76	1.47	1.05
382	Driver Journey Purpose=Commuting to/from work & Day of Week=Weekend	0.28	32.26	1.33	1.07
383	Driver Journey Purpose=Commuting to/from work & Day of Week=Weekend & Vehicle Location=Not at junction	0.14	34.89	1.44	1.08
384	Driver Journey Purpose=Commuting to/from work & Day of Week=Weekend & Junction Detail=Not at junction	0.14	34.89	1.44	1.08
385	Driver Journey Purpose=Commuting to/from work & Day of Week=Weekend & Junction Control=Not at junction or within 20 metres	0.14	34.64	1.43	1.07
386	Driver Journey Purpose=Commuting to/from work & Vehicle Skidding and Overturning=Yes	0.13	32.14	1.32	1.07
387	Driver Journey Purpose=Commuting to/from work & Vehicle Skidding and Overturning=Yes & Weather=Fine no high winds	0.10	34.33	1.41	1.07
388	Driver Journey Purpose=Commuting to/from work & Vehicle Skidding and Overturning=Yes & Vehicle 1st Point of Impact=Front	0.11	33.93	1.40	1.06
389	Driver Journey Purpose=Commuting to/from work & Pedestrian Crossing Physical Facilities=Central refuge	0.15	31.99	1.32	1.06
390	Driver Journey Purpose=Commuting to/from work & Pedestrian Crossing Physical Facilities=Central refuge & Pavement=Dry	0.12	34.33	1.41	1.07
391	Driver Journey Purpose=Commuting to/from work & Pedestrian Crossing Physical Facilities=Central refuge & Lighting=Daylight	0.11	33.77	1.39	1.06
392	Vehicle Towing and Articulation=Other	0.12	29.75	1.22	n.a.
393	Vehicle Towing and Articulation=Other & Weather=Fine no high winds	0.10	31.48	1.30	1.06
394	Driver Home Area=Small town	1.37	29.37	1.21	n.a.
395	Driver Home Area=Small town & Speed limit≥50	0.14	36.95	1.52	1.26
396	Driver Home Area=Small town & Speed limit≥50 & Driver Gender=M	0.11	39.66	1.63	1.07
397	Driver Home Area=Small town & Speed limit≥50 & Vehicle Manoeuvre=Going ahead	0.12	38.92	1.60	1.05
398	Driver Home Area=Small town & Vehicle Engine Capacity ≤1000	0.13	35.66	1.47	1.21
399	Driver Home Area=Small town & Driver Age≥75	0.13	35.57	1.46	1.21
400	Driver Home Area=Small town & Lighting=Darkness - lights lit	0.29	33.28	1.37	1.13
401	Driver Home Area=Small town & Lighting=Darkness - lights lit & Pavement=Wet or damp	0.16	36.49	1.50	1.10
402	Driver Home Area=Small town & Lighting=Darkness - lights lit & Vehicle 1st Point of Impact=Front	0.22	36.41	1.50	1.09
403	Driver Home Area=Small town & Lighting=Darkness - lights lit & Vehicle Manoeuvre=Going ahead	0.20	35.28	1.45	1.06

ID rule	Rules with driver characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
404	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work	0.18	33.06	1.36	1.13
405	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Driver Gender=M	0.14	39.48	1.63	1.19
406	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Vehicle Location=Not at junction	0.11	37.88	1.56	1.15
407	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Junction Detail=Not at junction	0.11	37.88	1.56	1.15
408	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Junction Control=Not at junction or within 20 metres	0.11	37.88	1.56	1.15
409	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Vehicle Manoeuvre=Going ahead	0.13	37.24	1.53	1.13
410	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.13	35.51	1.46	1.07
411	Driver Home Area=Small town & Driver Journey Purpose=Commuting to/from work & Road Type=Single carriageway	0.16	34.95	1.44	1.06
412	Driver Home Area=Small town & Driver IMD Decile=More deprived	0.40	31.72	1.31	1.08
413	Driver Home Area=Small town & Driver IMD Decile=More deprived & Driver Gender=M	0.29	33.63	1.38	1.06
414	Driver Home Area=Small town & Driver IMD Decile=More deprived & Vehicle Engine Capacity =1000-1500	0.11	33.33	1.37	1.05
415	Driver Home Area=Small town & 1st Road Class=B	0.20	31.70	1.31	1.08
416	Driver Home Area=Small town & 1st Road Class=B & Driver Gender=M	0.14	36.88	1.52	1.16
417	Driver Home Area=Small town & 1st Road Class=B & Vehicle Location=At junction	0.11	35.64	1.47	1.12
418	Driver Home Area=Small town & 1st Road Class=B & Vehicle 1st Point of Impact=Front	0.13	34.66	1.43	1.09
419	Driver Home Area=Small town & Pavement=Wet or damp	0.37	31.64	1.30	1.08
420	Driver Home Area=Small town & Pavement=Wet or damp & Vehicle Propulsion Code=Petrol	0.18	33.89	1.40	1.07
421	Driver Home Area=Small town & Pavement=Wet or damp & Vehicle Manoeuvre=Going ahead	0.23	33.33	1.37	1.05
422	Driver Home Area=Small town & Driver Age=0-24	0.20	31.52	1.30	1.07
423	Driver Home Area=Small town & Driver Age=0-24 & Vehicle 1st Point of Impact=Front	0.14	34.34	1.41	1.09
424	Driver Home Area=Small town & Vehicle Engine Capacity =Missing	0.14	31.02	1.28	1.06
425	Driver Home Area=Small town & Vehicle Engine Capacity =Missing & Speed limit=30	0.11	33.93	1.40	1.09
426	Driver Home Area=Small town & Vehicle Engine Capacity =Missing & Road Type=Single carriageway	0.14	33.83	1.39	1.09
427	Driver Home Area=Small town & Vehicle Engine Capacity =Missing & Driver Gender=M	0.12	33.06	1.36	1.07
428	Driver Age=0-24	3.06	29.32	1.21	n.a.
429	Driver Age=0-24 & Vehicle Skidding and Overturning=Yes	0.27	40.95	1.69	1.40
430	Driver Age=0-24 & Vehicle Skidding and Overturning=Yes & Junction Control=Not at junction or within 20 metres	0.14	44.09	1.82	1.08
431	Driver Age=0-24 & Vehicle Skidding and Overturning=Yes & Vehicle Location=Not at junction	0.14	43.84	1.80	1.07
432	Driver Age=0-24 & Vehicle Skidding and Overturning=Yes & Junction Detail=Not at junction	0.14	43.84	1.80	1.07
433	Driver Age=0-24 & Vehicle Skidding and Overturning=Yes & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.18	43.31	1.78	1.06
434	Driver Age=0-24 & Vehicle Skidding and Overturning=Yes & Vehicle Type=Car	0.16	43.03	1.77	1.05
435	Driver Age=0-24 & Speed limit≥50	0.14	38.56	1.59	1.31
436	Driver Age=0-24 & Speed limit≥50 & Vehicle 1st Point of Impact=Front	0.10	41.72	1.72	1.08
437	Driver Age=0-24 & Speed limit=40	0.16	38.16	1.57	1.30
438	Driver Age=0-24 & Speed limit=40 & 1st Road Class=A	0.12	41.29	1.70	1.08
439	Driver Age=0-24 & Speed limit=40 & Vehicle Manoeuvre=Going ahead	0.14	40.27	1.66	1.06
440	Driver Age=0-24 & Road Type=Dual carriageway	0.35	35.16	1.45	1.20

ID rule	Rules with driver characteristics as first antecedent and serious injury crashes as consequent Antecents	S %	C %	Lift	LIC
441	Driver Age=0-24 & Road Type=Dual carriageway & Lighting=Darkness - lights lit	0.14	37.70	1.55	1.07
442	Driver Age=0-24 & Road Type=Dual carriageway & Vehicle Manoeuvre=Going ahead	0.25	37.20	1.53	1.06
443	Driver Age=0-24 & Junction Detail=Other junction	0.16	33.44	1.38	1.14
444	Driver Age=0-24 & Junction Detail=Other junction & Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.10	35.98	1.48	1.08
445	Driver Age=0-24 & Junction Detail=Other junction & Vehicle Type=Car	0.13	35.80	1.47	1.07
446	Driver Age=0-24 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing	0.45	32.94	1.36	1.12
447	Driver Age=0-24 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing & Day of Week=Weekend	0.13	38.70	1.59	1.17
448	Driver Age=0-24 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing & Vehicle 1st Point of Impact=Nearside/Offside	0.12	38.05	1.57	1.16
449	Driver Age=0-24 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing & Junction Detail=T or staggered junction	0.16	36.18	1.49	1.10
450	Driver Age=0-24 & Pedestrian Crossing Physical Facilities=Pelican/puffin/toucan or similar non-junction pedestrian light crossing & Vehicle Manoeuvre=Going ahead	0.34	34.91	1.44	1.06
451	Driver Age=0-24 & Area=Rural	0.40	32.46	1.34	1.11
452	Driver Age=0-24 & Area=Rural & Pavement=Wet or damp	0.13	40.36	1.66	1.24
453	Driver Age=0-24 & Area=Rural & 1st Road Class=A	0.16	38.08	1.57	1.17
454	Driver Age=0-24 & Area=Rural & Vehicle 1st Point of Impact=Front	0.27	35.74	1.47	1.10
455	Driver Age=0-24 & Area=Rural & Vehicle Location=At junction	0.14	35.56	1.46	1.10
456	Driver Age=0-24 & Area=Rural & Junction Control=Give way/uncontrolled	0.12	35.44	1.46	1.09
457	Driver Age=0-24 & Area=Rural & Driver Gender=M	0.28	34.50	1.42	1.06
458	Driver Age=0-24 & Number of Vehicles=2	0.22	31.89	1.31	1.09
459	Driver Age=0-24 & Number of Vehicles=2 & Vehicle Location=At junction	0.11	34.55	1.42	1.08
460	Driver Age=0-24 & Lighting=Darkness - lights lit	0.97	31.81	1.31	1.08
461	Driver Age=0-24 & Lighting=Darkness - lights lit & Vehicle Manoeuvre=Going ahead	0.70	34.26	1.41	1.08
462	Driver Age=0-24 & Lighting=Darkness - lights lit & Vehicle 1st Point of Impact=Front	0.72	33.52	1.38	1.05
463	Driver Age=0-24 & Vehicle Manoeuvre=Going ahead	2.18	31.75	1.31	1.08
464	Driver Age=0-24 & Vehicle Manoeuvre=Going ahead & Driver Journey Purpose=Commuting to/from work	0.30	35.68	1.47	1.12
465	Driver Age=0-24 & Vehicle Manoeuvre=Going ahead & Day of Week=Weekend	0.61	34.56	1.42	1.09
466	Driver Age=0-24 & Vehicle Manoeuvre=Going ahead & Junction Control=Auto traffic signal	0.26	33.66	1.39	1.06
467	Driver Age=0-24 & Vehicle Manoeuvre=Going ahead & Junction Detail=T or staggered junction	0.66	33.54	1.38	1.06
468	Driver Age=0-24 & Vehicle Manoeuvre=Going ahead & 1st Road Class=A	0.82	33.51	1.38	1.06
469	Driver Age=0-24 & Driver Journey Purpose=Commuting to/from work	0.40	31.61	1.30	1.08
470	Driver Age=0-24 & Driver Journey Purpose=Commuting to/from work & Junction Detail=T or staggered junction	0.14	34.89	1.44	1.10
471	Driver Age=0-24 & Driver Journey Purpose=Commuting to/from work & Driver Gender=M	0.27	33.27	1.37	1.05
472	Driver Age=0-24 & X2nd Road Class=C	0.10	31.22	1.29	1.06
473	Driver Age=0-24 & Day of Week=Weekend	0.81	31.21	1.29	1.06
474	Driver Age=0-24 & Day of Week=Weekend & Pavement=Wet or damp	0.21	33.10	1.36	1.06
475	Driver Age=0-24 & Day of Week=Weekend & Vehicle 1st Point of Impact=Front	0.57	32.90	1.35	1.05

Artificial neural network

Table 154 – Artificial Neural Network parameter estimates, Great Britain.

		Predicted												
		Hidden Layer 1												
Predictor		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)
Input Layer	(Bias)	-0.173	0.368	-0.380	0.119	0.412	0.340	-0.492	-0.376	-0.285	-0.236	0.174	0.299	0.154
	Number of Vehicles≥2	-0.428	0.353	0.332	-0.295	0.104	0.393	-0.322	0.159	-0.218	0.478	0.377	-0.198	-0.239
	Number of Vehicles=1	0.507	0.106	0.233	-0.126	-0.198	-0.088	-0.443	0.419	0.087	0.099	0.014	-0.344	0.428
	Number of Vehicles=2	-0.356	0.378	-0.450	0.463	-0.304	-0.498	0.143	-0.268	-0.121	-0.320	0.106	-0.232	0.234
	Day of Week=Weekday	0.018	0.299	-0.259	0.069	0.083	0.299	0.174	-0.518	0.244	-0.072	0.050	-0.211	0.182
	Day of Week=Weekend	-0.278	-0.117	0.299	-0.364	0.477	-0.114	0.104	-0.295	0.230	-0.335	-0.357	-0.199	-0.150
	1st Road Class=A	-0.027	0.388	-0.298	0.467	-0.039	0.312	-0.165	-0.042	0.292	-0.456	-0.004	0.011	-0.015
	1st Road Class=B	-0.413	0.233	0.120	-0.339	0.431	-0.247	-0.462	0.143	-0.007	0.118	-0.063	0.489	-0.163
	1st Road Class=C	-0.240	0.049	-0.312	-0.359	0.282	-0.321	0.145	-0.234	-0.117	0.395	0.162	0.382	-0.377
	1st Road Class=Motorway	-0.185	-0.296	0.486	-0.403	-0.450	0.004	-0.362	0.116	-0.295	0.139	0.005	0.185	0.026
	Road Type=Dual carriageway	-0.045	-0.284	-0.436	-0.457	0.465	0.310	0.297	0.228	-0.269	0.047	0.239	-0.439	0.074
	Road Type=One way street	-0.007	0.427	0.373	0.480	0.181	0.253	0.439	0.282	-0.089	0.003	-0.367	0.344	-0.030
	Road Type=Roundabout	-0.303	0.131	-0.296	0.406	0.020	0.198	0.423	0.222	-0.001	-0.153	-0.163	-0.218	-0.355
	Road Type=Single carriageway	0.321	-0.064	-0.035	-0.165	0.195	0.137	0.468	0.291	0.313	-0.002	0.030	-0.455	-0.130
	Road Type=Slip road	-0.425	0.427	-0.339	0.486	-0.416	0.269	-0.418	-0.048	-0.089	-0.270	0.235	0.401	0.447
	Speed limit≥50	-0.162	0.163	-0.380	0.489	-0.198	-0.339	-0.161	-0.321	0.185	-0.118	0.436	-0.022	0.342
	Speed limit=20	0.138	-0.155	-0.261	-0.452	-0.375	-0.124	0.460	0.112	0.071	0.274	-0.031	-0.259	0.112
	Speed limit=30	-0.120	0.221	0.028	-0.187	0.082	0.346	-0.292	0.078	-0.488	0.427	-0.228	-0.030	0.226
	Speed limit=40	0.047	0.358	0.306	0.033	-0.065	-0.038	0.019	0.149	0.487	0.133	0.290	0.201	-0.234
	Junction Detail=Crossroads	0.367	-0.003	-0.221	-0.222	-0.029	0.283	-0.273	0.249	-0.137	-0.159	-0.084	-0.362	-0.453
	Junction Detail=Mini-roundabout	0.206	-0.373	-0.419	-0.395	0.400	-0.317	0.305	0.174	0.115	0.359	0.111	0.105	-0.346
	Junction Detail=More than 4 arms (not roundabout)	0.202	0.224	-0.254	-0.421	-0.067	-0.416	-0.487	-0.474	0.354	0.493	-0.402	-0.336	0.107
	Junction Detail=Not at junction	0.185	-0.217	0.284	-0.031	-0.105	0.305	0.406	-0.165	-0.311	-0.321	-0.119	0.396	-0.326
	Junction Detail=Other junction	0.380	-0.347	0.314	-0.121	0.130	0.270	-0.373	0.289	0.289	0.432	0.191	-0.378	0.416
	Junction Detail=Private drive or entrance	0.501	-0.325	-0.074	0.007	0.195	-0.144	0.483	0.367	-0.269	0.410	0.082	0.323	0.486
	Junction Detail=Roundabout	0.426	0.173	0.256	-0.475	-0.240	-0.380	0.000	-0.197	-0.345	-0.308	-0.302	0.263	-0.464
	Junction Detail=Slip road	-0.263	0.436	0.242	-0.117	-0.466	0.206	-0.058	0.451	0.221	-0.099	0.445	-0.452	0.348
	Junction Detail=T or staggered junction	-0.382	0.158	0.347	0.286	-0.390	-0.474	0.081	-0.486	-0.498	-0.041	0.366	-0.478	-0.282
	Junction Control=Authorised person	0.154	0.070	0.461	-0.350	-0.345	-0.189	-0.291	-0.268	0.454	-0.327	0.119	0.431	0.072

Predictor	Predicted												
	Hidden Layer 1												
	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)
Junction Control=Auto traffic signal	0.347	0.401	-0.434	0.437	0.292	0.461	-0.180	-0.501	0.355	0.276	0.120	0.357	-0.253
Junction Control=Give way/uncontrolled	-0.359	0.099	-0.091	-0.360	-0.155	-0.154	0.129	0.016	-0.141	-0.166	0.372	0.502	0.045
Junction Control=Not at junction or within 20 metres	0.249	0.092	-0.010	-0.310	0.204	-0.445	0.095	0.033	-0.249	-0.236	-0.254	0.244	0.381
Junction Control=Stop sign	-0.213	-0.080	-0.242	-0.469	-0.151	-0.450	-0.444	-0.186	0.443	0.004	0.109	0.477	0.120
Pedestrian Crossing Human Control=None within 50 metres	0.077	0.161	0.466	0.122	-0.246	0.259	0.323	-0.061	-0.432	-0.003	0.432	0.134	0.092
Pedestrian Crossing Human Control=School crossing patrol	0.161	0.216	-0.419	0.065	-0.085	-0.223	-0.233	-0.409	0.254	0.200	-0.239	-0.006	0.220
Pedestrian Crossing Physical Facilities=Central refuge	0.047	0.299	-0.160	0.301	0.014	0.459	0.316	-0.256	-0.185	-0.183	-0.488	0.101	-0.490
Pedestrian Crossing Physical Facilities=Footbridge/subway	0.015	-0.278	-0.054	-0.206	-0.328	-0.024	0.290	0.497	0.438	0.267	-0.237	0.058	-0.234
Pedestrian Crossing Physical Facilities=No physical crossing facilities within 50 metres	0.153	-0.345	-0.061	-0.004	-0.459	-0.441	-0.432	-0.318	0.316	0.295	-0.228	0.134	0.467
Pedestrian Crossing Physical Facilities=Pedestrian phase at traffic signal junction	0.445	-0.140	-0.123	-0.320	-0.010	-0.264	0.139	-0.482	-0.283	-0.079	0.099	-0.400	-0.415
Pedestrian Crossing Physical Facilities=Pelican, puffin, toucan or similar non-junction pedestrian light crossing	0.055	-0.410	0.170	0.401	0.114	0.424	0.259	-0.165	-0.071	-0.428	-0.242	-0.131	-0.436
Pedestrian Crossing Physical Facilities=Zebra	-0.451	0.446	-0.218	0.229	-0.463	0.047	0.393	-0.434	-0.445	-0.077	0.244	-0.365	-0.023
Lighting=Darkness - lighting unknown	0.219	0.361	0.331	0.340	0.322	0.281	-0.291	0.248	-0.039	-0.238	-0.093	0.164	-0.467
Lighting=Darkness - lights lit	-0.093	-0.229	-0.074	0.381	-0.321	-0.072	0.403	-0.429	0.452	0.189	0.232	-0.398	0.008
Lighting=Darkness - lights unlit	-0.312	-0.200	-0.104	0.190	-0.311	-0.395	-0.325	0.418	-0.502	0.182	-0.297	0.246	0.406
Lighting=Darkness - no lighting	0.135	-0.452	0.362	-0.468	0.158	0.398	0.055	-0.283	0.488	0.118	-0.267	-0.109	0.304
Lighting=Daylight	0.473	-0.020	0.353	-0.460	-0.497	-0.047	0.049	0.190	0.158	0.154	-0.483	0.176	0.444
Weather=Fine + high winds	-0.145	0.062	0.099	0.368	-0.302	0.353	0.069	0.190	0.221	0.354	-0.133	0.016	0.341
Weather=Fine no high winds	0.460	0.327	0.093	0.006	-0.025	0.263	-0.349	0.447	0.217	-0.334	0.141	0.304	-0.089
Weather=Fog or mist	-0.042	-0.307	0.338	0.417	0.003	-0.253	-0.328	-0.071	-0.370	-0.008	-0.470	0.419	0.111
Weather=Raining + high winds	0.336	-0.348	0.291	0.117	-0.453	-0.211	0.487	-0.242	-0.407	-0.031	0.436	-0.181	0.168
Weather=Raining no high winds	0.077	-0.004	-0.002	-0.246	0.163	-0.025	-0.359	0.432	-0.094	0.053	-0.069	0.283	-0.224
Weather=Snowing	0.225	-0.268	-0.054	-0.188	0.332	0.090	0.052	-0.388	0.214	0.238	0.140	0.335	-0.079
Pavement=Dry	-0.338	0.581	-0.029	0.305	0.357	-0.084	0.463	-0.403	-0.577	-0.233	-0.259	-0.330	0.390
Pavement=Snowy/Frozen	-0.478	0.155	0.124	-0.078	-0.362	0.359	0.159	0.123	-0.372	0.008	0.389	-0.338	0.306
Pavement=Wet or damp	-0.288	0.344	-0.257	0.142	-0.326	0.147	0.022	0.393	0.393	0.035	-0.080	-0.056	-0.391
Area=Rural	-0.247	-0.238	0.307	0.491	0.293	-0.318	0.401	0.156	0.215	-0.231	0.347	-0.498	0.074
Area=Urban	0.066	-0.327	0.089	0.365	0.415	0.267	-0.323	-0.197	-0.376	0.519	-0.365	0.121	0.185
Vehicle Type=Bicycle	0.098	-0.432	0.107	-0.109	-0.141	0.423	0.282	0.466	-0.424	0.273	-0.403	-0.390	0.493

	Predicted												
	Hidden Layer 1												
Predictor	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)
Vehicle Type=Bus	-0.417	0.167	0.300	0.373	-0.240	0.323	0.112	0.223	-0.289	-0.151	-0.050	-0.219	-0.067
Vehicle Type=Car	-0.350	0.614	0.020	-0.287	-0.125	0.312	0.492	0.115	0.296	0.578	-0.157	-0.167	-0.043
Vehicle Type=PTW<500	0.371	-0.023	-0.235	0.246	-0.429	0.154	-0.351	0.096	-0.382	-0.363	-0.016	0.345	0.472
Vehicle Type=PTW≥500	-0.127	0.300	0.423	0.320	0.053	-0.159	-0.257	-0.038	-0.127	-0.257	-0.429	-0.018	0.096
Vehicle Type=Truck	0.207	0.002	0.270	-0.033	0.560	0.444	0.001	0.385	0.381	0.227	-0.168	-0.261	0.349
Vehicle Type=Van	-0.168	-0.075	-0.066	-0.058	0.118	-0.126	-0.175	-0.168	-0.481	0.161	0.164	0.314	0.448
Vehicle Towing and Articulation=Articulated vehicle	0.305	0.353	0.014	-0.075	-0.389	-0.215	-0.138	0.383	0.183	0.441	0.342	0.295	0.247
Vehicle Towing and Articulation=No tow/articulation	0.371	-0.196	-0.109	0.063	-0.507	0.123	0.400	-0.496	0.376	-0.250	-0.569	-0.461	-0.448
Vehicle Manoeuvre=Going ahead	-0.344	0.318	-0.453	0.065	0.286	0.329	0.057	-0.283	-0.244	-0.140	0.358	0.434	0.080
Vehicle Manoeuvre=Moving off	-0.081	0.259	0.251	-0.251	-0.105	0.284	-0.154	0.456	0.240	-0.178	0.118	-0.470	0.458
Vehicle Manoeuvre=Overtaking	0.204	0.164	-0.287	-0.131	0.478	0.341	-0.082	0.210	0.071	0.057	0.055	0.348	-0.123
Vehicle Manoeuvre=Reversing	-0.236	-0.127	-0.433	-0.150	-0.482	0.312	0.402	-0.077	-0.165	-0.176	-0.350	0.124	-0.352
Vehicle Manoeuvre=Turning left/right/U	0.490	-0.005	-0.217	0.174	-0.241	0.183	0.346	-0.463	-0.310	-0.440	-0.202	-0.472	0.309
Vehicle Junction Location=At junction	-0.064	0.169	-0.042	-0.454	-0.262	0.079	-0.159	0.025	0.101	0.411	0.065	-0.363	-0.351
Vehicle Junction Location=Not at junction	-0.241	-0.193	-0.290	0.157	0.138	-0.408	-0.073	0.019	0.138	0.219	0.043	-0.490	0.428
Vehicle Skidding and Overturning=No	0.214	-0.219	0.337	-0.295	-0.430	-0.475	-0.193	-0.091	0.152	0.008	-0.384	0.421	-0.210
Vehicle Skidding and Overturning=Yes	0.309	0.179	-0.133	-0.444	-0.400	0.334	-0.106	-0.247	0.330	-0.372	0.241	0.107	0.232
Vehicle 1st Point of Impact=Back	0.044	0.158	-0.315	-0.411	0.294	-0.440	0.136	0.079	-0.023	-0.136	0.079	0.100	0.126
Vehicle 1st Point of Impact=Front	0.087	0.447	0.346	-0.052	0.520	-0.226	-0.484	-0.052	0.312	0.090	0.254	-0.504	0.401
Vehicle 1st Point of Impact=Nearside/Offside	0.392	-0.305	-0.386	0.168	-0.479	-0.482	-0.032	0.175	0.180	0.429	-0.093	0.358	0.368
Vehicle 1st Point of Impact=No impact	-0.421	-0.012	0.464	-0.496	0.218	0.323	0.409	-0.001	-0.040	0.495	-0.073	0.003	-0.099
Driver Gender=F	0.131	0.117	-0.049	-0.443	-0.043	0.141	0.316	0.030	-0.363	-0.157	-0.515	0.138	-0.316
Driver Gender=M	0.232	-0.001	-0.415	-0.482	-0.214	0.132	-0.279	0.065	-0.196	-0.176	0.320	0.142	0.158
Driver Age≥75	0.375	0.263	0.172	0.237	0.382	-0.190	0.258	0.372	0.451	-0.468	-0.433	-0.216	-0.418
Driver Age=0-24	-0.376	-0.152	-0.297	-0.360	0.219	-0.155	-0.092	0.329	-0.255	-0.432	0.409	-0.402	0.193
Driver Age=25-34	0.068	-0.389	-0.219	-0.224	0.326	-0.023	-0.194	0.411	-0.297	-0.169	0.037	0.448	0.438
Driver Age=35-44	-0.164	-0.109	-0.450	-0.122	0.225	-0.099	-0.409	-0.320	-0.497	-0.251	0.453	0.117	0.332
Driver Age=45-54	0.088	-0.090	-0.489	0.339	-0.079	0.479	-0.023	0.015	-0.250	-0.490	0.203	-0.476	0.489
Driver Age=55-64	0.415	0.397	-0.402	-0.338	0.399	-0.494	-0.199	-0.011	0.460	0.072	0.251	-0.196	0.301
Driver Age=65-74	-0.182	-0.140	-0.458	0.169	-0.455	0.058	-0.474	0.036	0.377	0.146	-0.315	-0.323	0.108
Vehicle Engine Capacity≤1000	0.152	0.407	-0.372	0.146	0.121	-0.143	0.454	0.379	-0.245	0.283	-0.364	0.159	0.450
Vehicle Engine Capacity≥3000	0.359	-0.292	-0.450	0.416	0.277	-0.032	-0.101	0.378	0.289	-0.458	-0.371	-0.273	0.184

Predictor	Predicted												
	Hidden Layer 1												
	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)
Vehicle Engine Capacity=1000-1500	0.307	0.347	0.137	-0.472	-0.014	-0.084	-0.068	0.164	0.386	0.499	-0.373	0.320	-0.035
Vehicle Engine Capacity=1500-2000	-0.203	0.183	0.462	-0.059	0.253	0.055	-0.337	0.264	-0.025	0.447	0.264	-0.154	-0.440
Vehicle Engine Capacity=2000-3000	-0.353	0.049	0.318	-0.298	-0.355	-0.217	0.386	0.202	-0.145	-0.295	0.459	0.388	-0.345
Vehicle Propulsion Code=Heavy oil	0.113	-0.087	0.317	0.489	-0.187	0.140	0.250	-0.131	0.310	0.231	0.467	-0.054	-0.446
Vehicle Propulsion Code=Hybrid electric	0.363	-0.442	-0.288	0.045	0.111	-0.122	0.199	-0.351	-0.356	-0.439	-0.401	0.030	-0.471
Vehicle Propulsion Code=Petrol	-0.385	0.201	-0.062	0.210	-0.071	0.288	-0.259	-0.183	-0.004	-0.046	-0.230	0.285	-0.307
Vehicle Age>15	-0.176	0.189	-0.465	0.004	-0.228	0.087	0.101	-0.431	-0.151	0.286	0.297	0.114	0.245
Vehicle Age≤15	0.052	0.007	-0.130	0.388	0.421	-0.239	0.299	-0.275	0.270	-0.193	-0.130	0.379	-0.294
Pedestrian1 Gender=F	-0.342	0.095	-0.018	0.241	0.044	-0.329	0.043	-0.133	-0.113	0.419	-0.106	0.173	0.301
Pedestrian1 Gender=M	0.074	-0.236	-0.024	-0.174	0.078	-0.412	-0.027	-0.148	-0.247	0.225	0.448	-0.377	-0.220
Pedestrian1 Age≥75	-0.114	0.234	-0.366	-0.213	0.114	-0.176	-0.396	-0.015	-0.308	-0.603	0.475	-0.437	-0.416
Pedestrian1 Age=0-14	0.087	0.498	-0.385	-0.079	0.359	0.444	0.151	0.293	-0.182	-0.227	-0.277	-0.161	0.207
Pedestrian1 Age=15-24	0.090	-0.143	-0.123	-0.456	-0.012	-0.431	-0.125	-0.089	0.422	0.277	-0.392	-0.310	0.400
Pedestrian1 Age=25-34	0.285	-0.451	-0.093	0.324	0.051	-0.064	-0.024	-0.355	-0.510	-0.177	0.131	-0.319	-0.051
Pedestrian1 Age=35-44	-0.458	-0.252	-0.482	0.443	-0.025	0.193	-0.264	-0.174	0.158	-0.247	0.079	0.346	0.219
Pedestrian1 Age=45-54	0.113	-0.283	0.191	-0.064	-0.063	-0.236	0.254	0.314	-0.194	0.453	0.297	-0.126	0.163
Pedestrian1 Age=55-64	-0.032	0.036	0.140	0.301	0.312	-0.460	-0.038	-0.501	0.031	-0.005	-0.368	0.468	0.320
Pedestrian1 Age=65-74	0.341	0.177	0.420	-0.490	0.418	-0.102	-0.441	0.215	0.347	-0.209	-0.470	-0.273	0.069

Table 155 – Artificial Neural Network parameter estimates for the output layer, Great Britain.

Predictor		Predicted		
		Output Layer		
		Crash Severity=Fatal	Crash Severity=Serious	Crash Severity=Slight
Hidden Layer 1	(Bias)	0.354	0.014	0.071
	H(1:1)	0.076	0.067	0.463
	H(1:2)	-0.066	0.289	0.127
	H(1:3)	-0.251	-0.131	-0.074
	H(1:4)	0.035	-0.164	-0.116
	H(1:5)	0.400	-0.007	-0.274
	H(1:6)	-0.244	-0.121	-0.149
	H(1:7)	-0.014	0.197	0.443
	H(1:8)	0.057	-0.147	0.036
	H(1:9)	0.290	-0.196	-0.288
	H(1:10)	-0.450	0.460	0.448
	H(1:11)	0.606	0.119	-0.205
	H(1:12)	-0.140	-0.224	-0.007
	H(1:13)	-0.264	-0.153	0.070

APPENDIX 2 ~ SWEDEN

Classification tree

Table 156 – Tree in table format, Sweden.

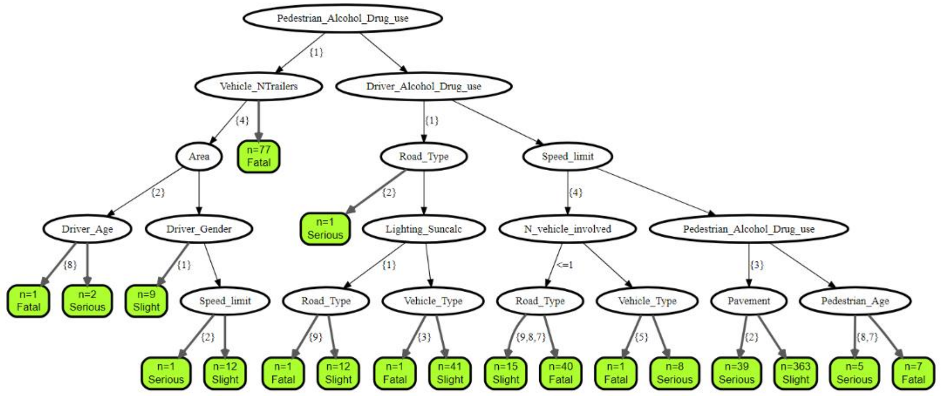
Node	Fatal		Serious		Slight		Total		Predicted Category	Parent Node
	N	%	N	%	N	%	N	%		
0	212	2.2	426	4.5	8,788	93.2	9,426	100.0	Slight	
1	74	7.2	53	5.2	896	87.6	1,023	19.2	Fatal	0
2	138	1.6	373	4.4	7,892	93.9	8,403	80.8	Slight	0
3	59	14.2	29	7.0	327	78.8	415	12.8	Fatal	1
4	15	2.5	24	3.9	569	93.6	608	6.4	Fatal	1
5	92	3.5	214	8.2	2,290	88.2	2,596	39.9	Serious	2
6	46	.8	159	2.7	5,602	96.5	5,807	40.9	Slight	2
7	0	0.0	3	1.4	209	98.6	212	1.0	Slight	4
8	15	3.8	21	5.3	360	90.9	396	5.4	Fatal	4
9	0	0.0	30	7.0	401	93.0	431	3.9	Serious	5
10	92	4.2	184	8.5	1,889	87.3	2,165	36.0	Fatal	5
11	25	.5	132	2.6	4,896	96.9	5,053	32.8	Slight	6
12	21	2.8	27	3.6	706	93.6	754	8.1	Fatal	6
13	7	2.3	18	5.9	282	91.9	307	3.6	Serious	8
14	8	9.0	3	3.4	78	87.6	89	1.8	Fatal	8
15	8	12.9	2	3.2	52	83.9	62	1.6	Fatal	10
16	84	4.0	182	8.7	1,837	87.4	2,103	34.4	Serious	10
17	4	1.5	20	7.7	236	90.8	260	3.1	Serious	11
18	21	.4	112	2.3	4,660	97.2	4,793	29.7	Slight	11
19	9	1.5	21	3.4	581	95.1	611	5.3	Slight	12
20	12	8.4	6	4.2	125	87.4	143	2.8	Fatal	12

Table 157 – Posterior Classification Ratio (PCR) for all nodes, Sweden.

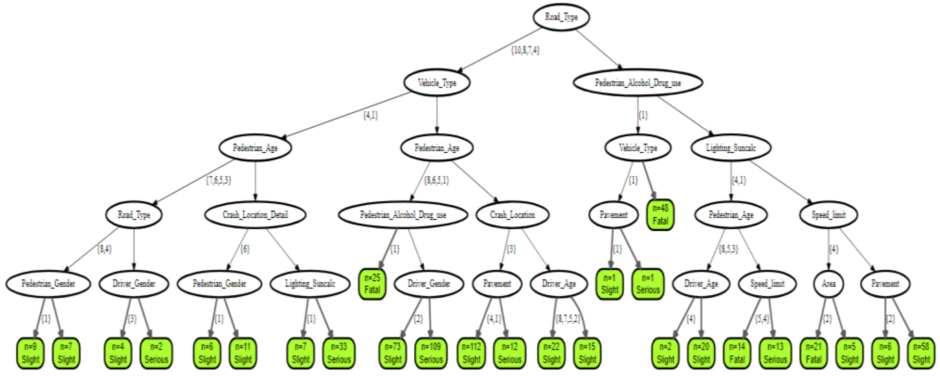
Node	PCR			
	Fatal	Serious	Slight	Actual Predicted Class
0	1.00	1.00	1.00	-
1	3.22	1.15	0.94	Fatal
2	0.73	0.98	1.01	Slight
3	6.32	1.55	0.85	Fatal
4	1.10	0.87	1.00	Fatal
5	1.58	1.82	0.95	Serious
6	0.35	0.61	1.03	Slight
7	0.00	0.31	1.06	Slight
8	1.68	1.17	0.98	Fatal
9	0.00	1.54	1.00	Serious
10	1.89	1.88	0.94	Fatal
11	0.22	0.58	1.04	Slight
12	1.24	0.79	1.00	Fatal
13	1.01	1.30	0.99	Serious
14	4.00	0.75	0.94	Fatal
15	5.74	0.71	0.90	Fatal
16	1.78	1.91	0.94	Serious
17	0.68	1.70	0.97	Serious
18	0.19	0.52	1.04	Slight
19	0.65	0.76	1.02	Slight
20	3.73	0.93	0.94	Fatal

Random forest

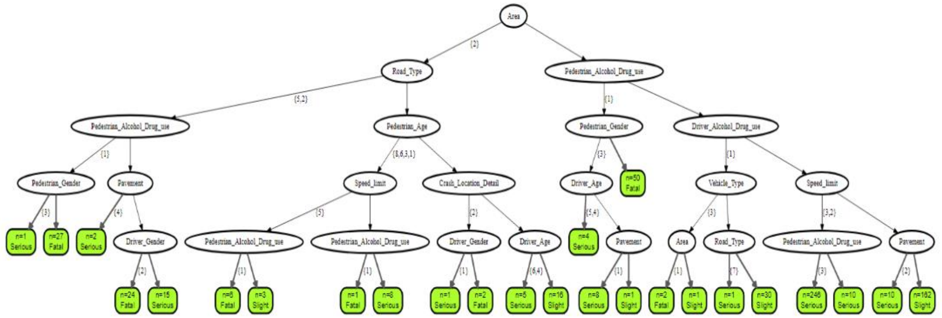
1



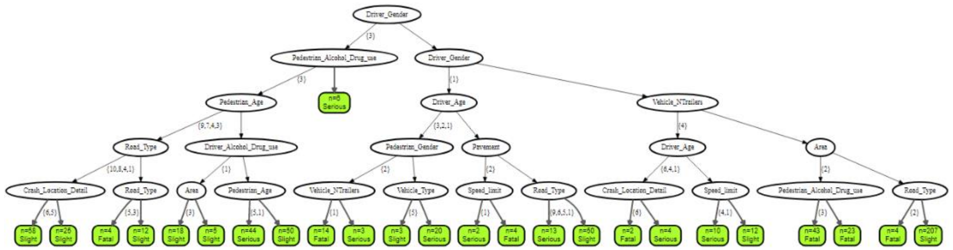
2



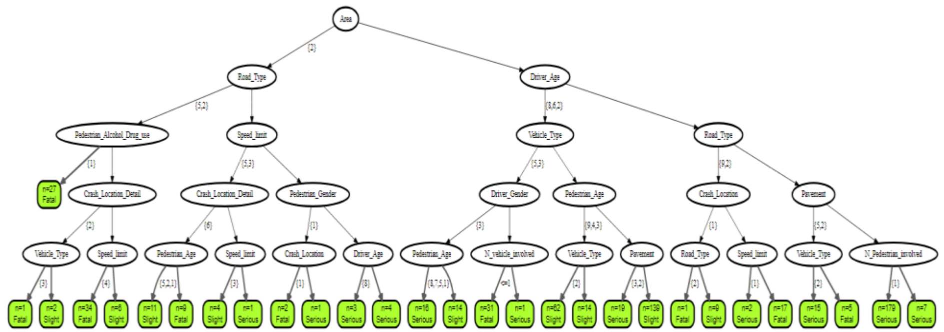
3



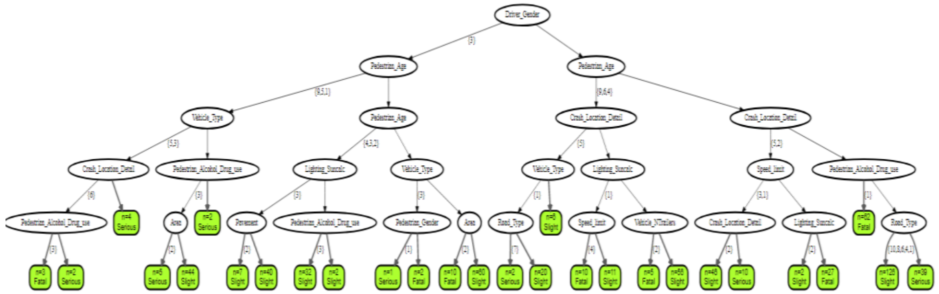
4



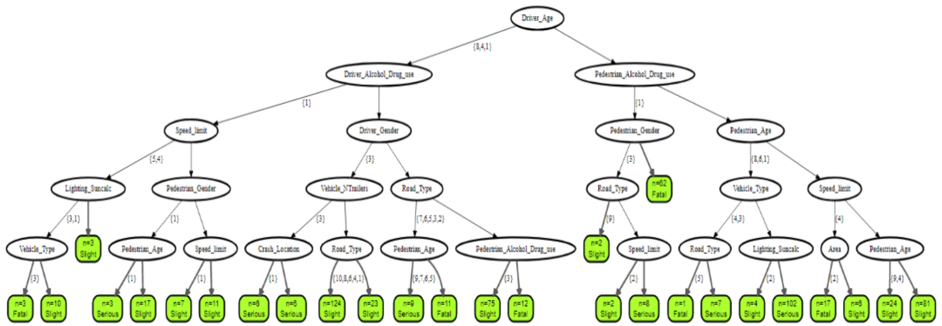
5



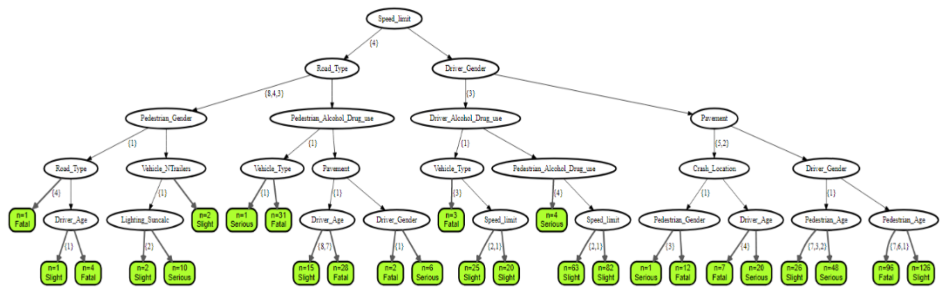
6



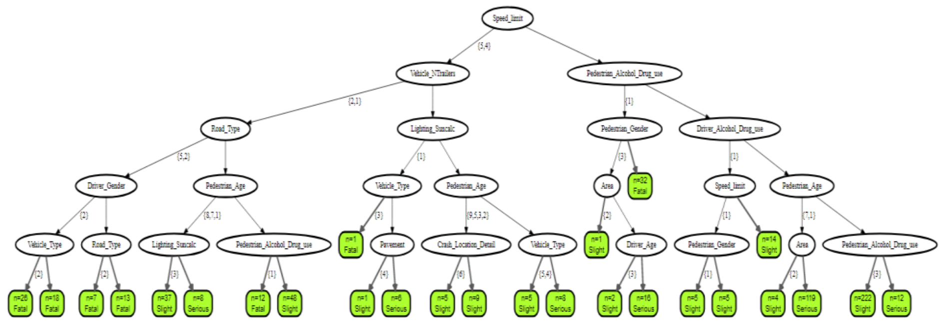
7



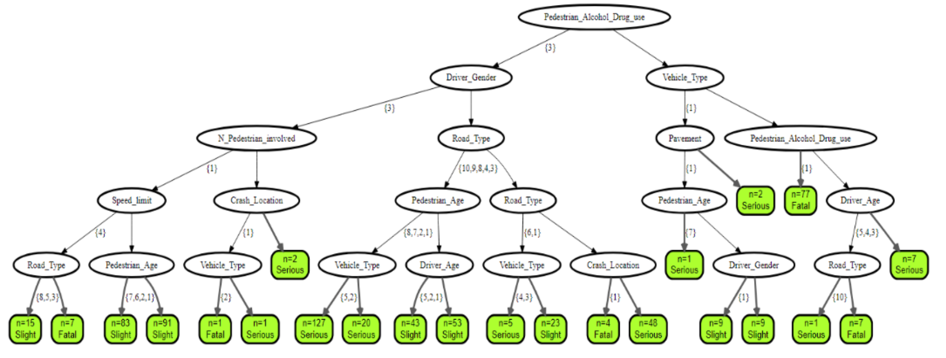
8



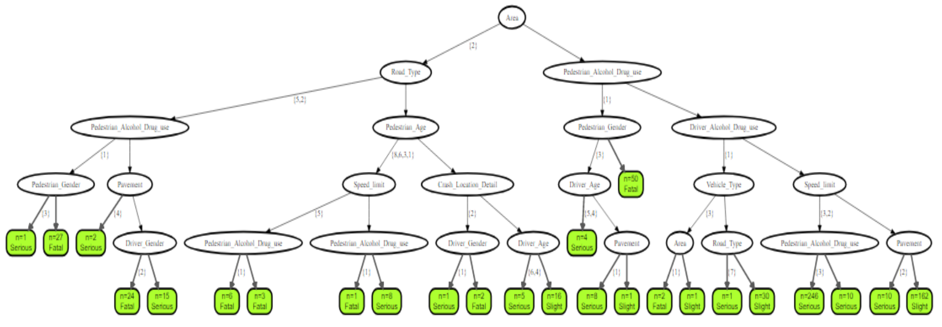
9



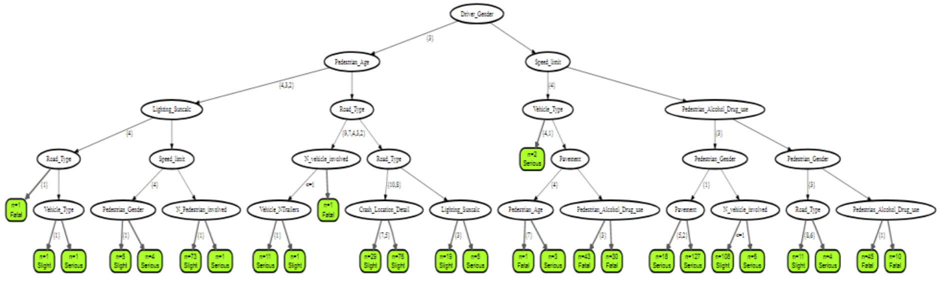
10



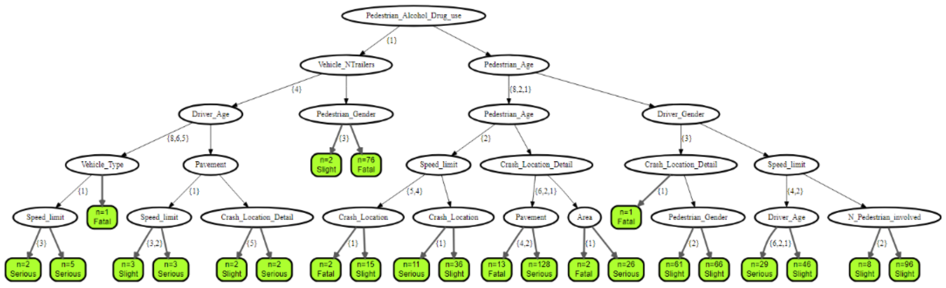
11



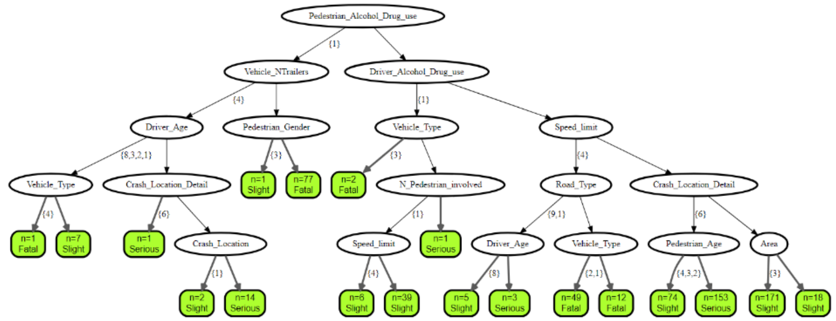
12



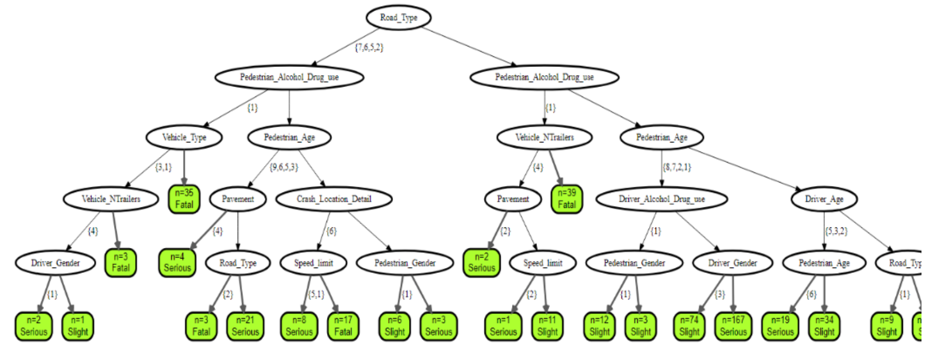
13



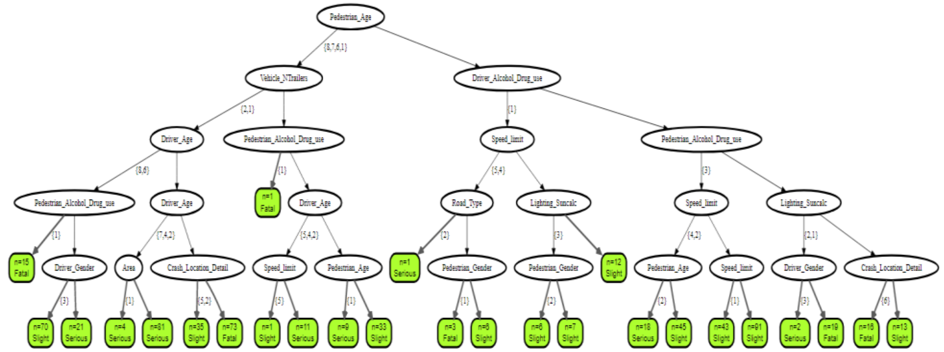
14



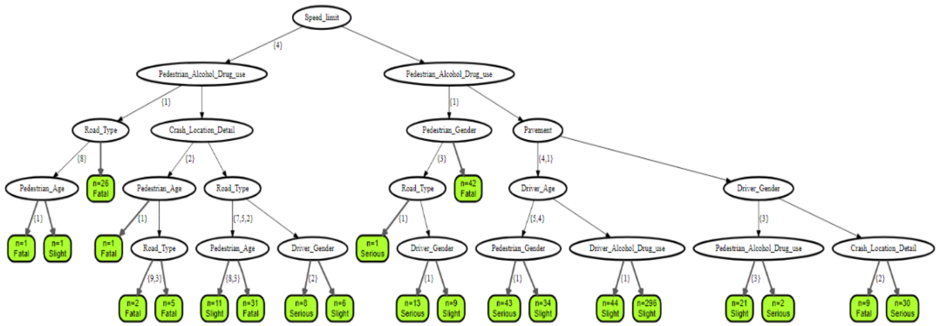
15



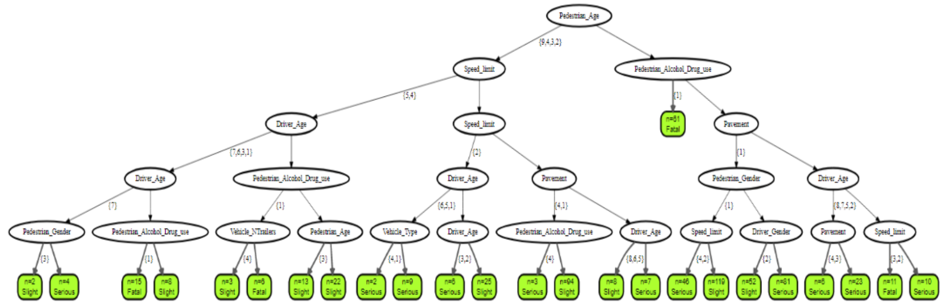
16



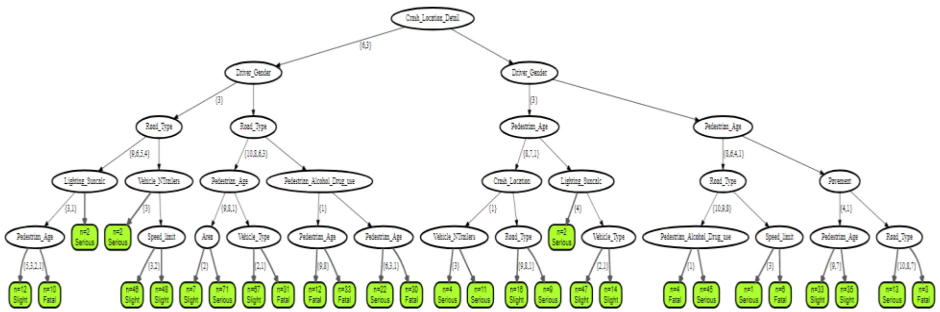
17



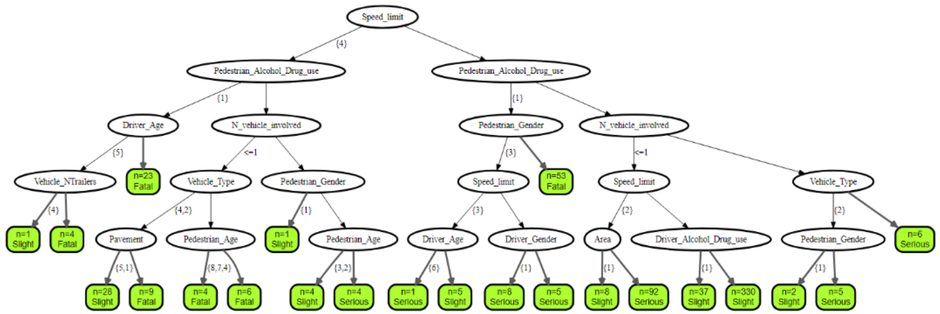
18



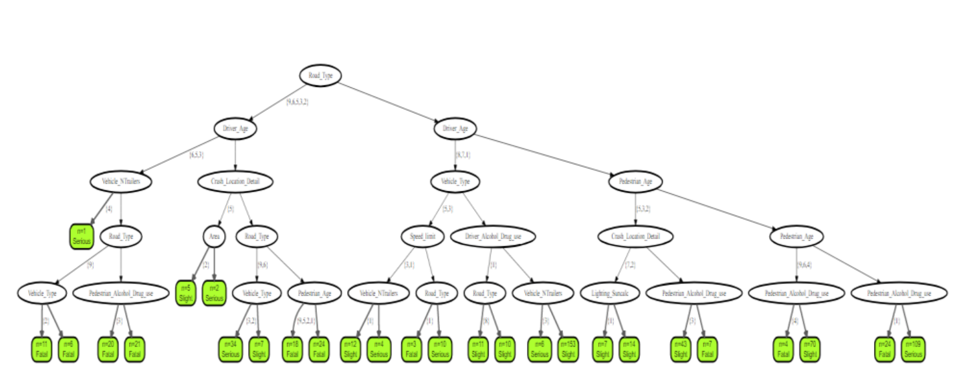
19



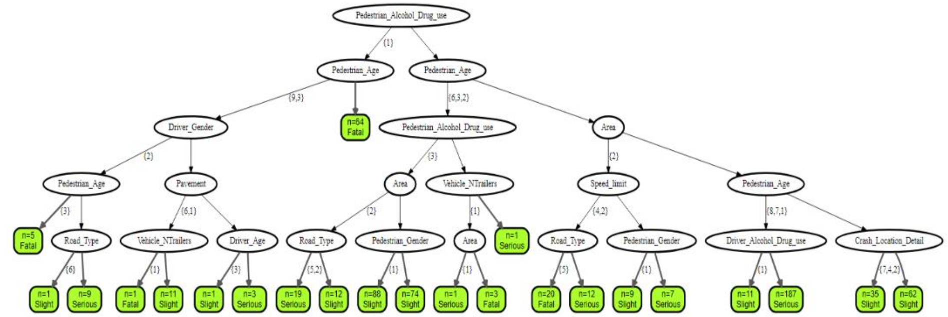
20



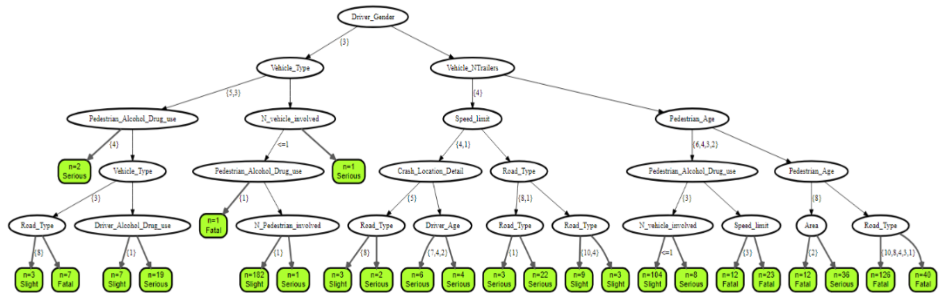
21



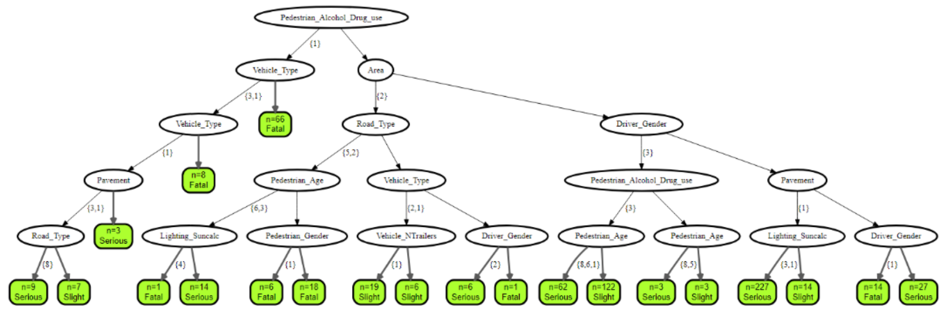
22


Pedestrian_Alcohol_Drug_use
Pedestrian_Age
Driver_Gender
Pavement
Area
Vehicle_NTrailers
Speed_limit
Road_Type
Driver_Age
Pedestrian_Gender
Driver_Alcohol_Drug_use
Crash_Location_Detail

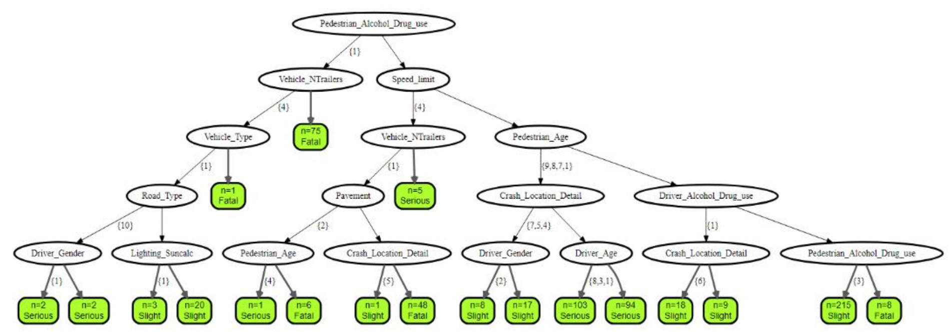
23


Driver_Gender
Vehicle_Type
Vehicle_NTrailers
Pedestrian_Alcohol_Drug_use
N_vehicle_involved
Speed_limit
Pedestrian_Age
Crash_Location_Detail
Road_Type
Driver_Alcohol_Drug_use
N_Pedestrian_involved
Driver_Age
Area

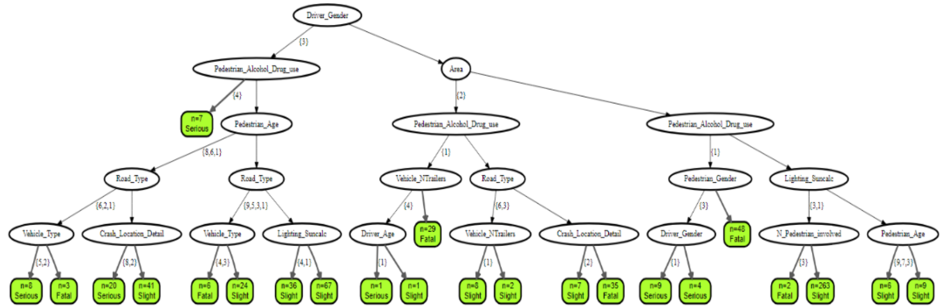
24


Pedestrian_Alcohol_Drug_use
Vehicle_Type
Area
Road_Type
Driver_Gender
Pavement
Pedestrian_Age
Lighting_Suncalc
Pedestrian_Gender
Vehicle_NTrailers

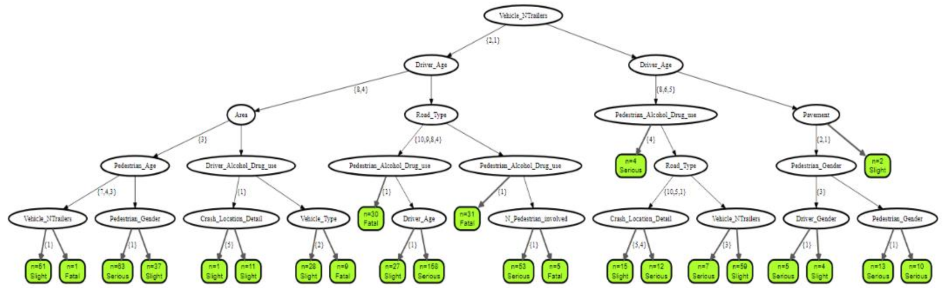
25



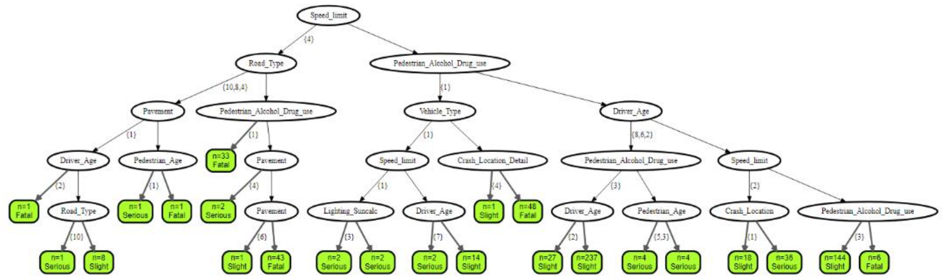
26



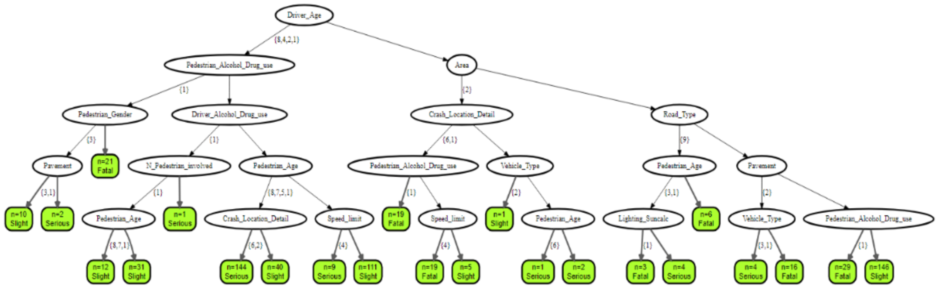
27



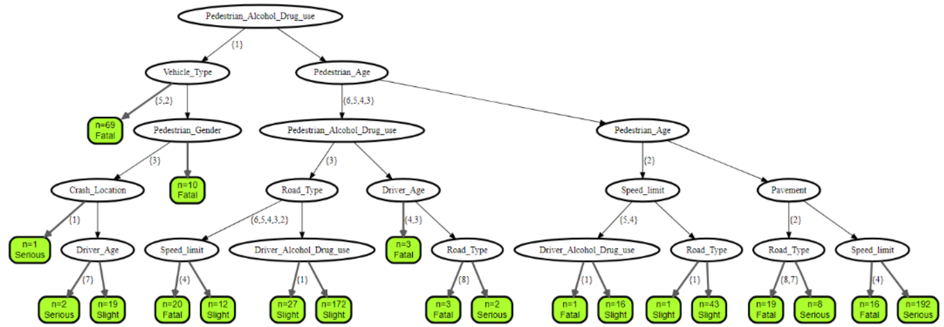
28



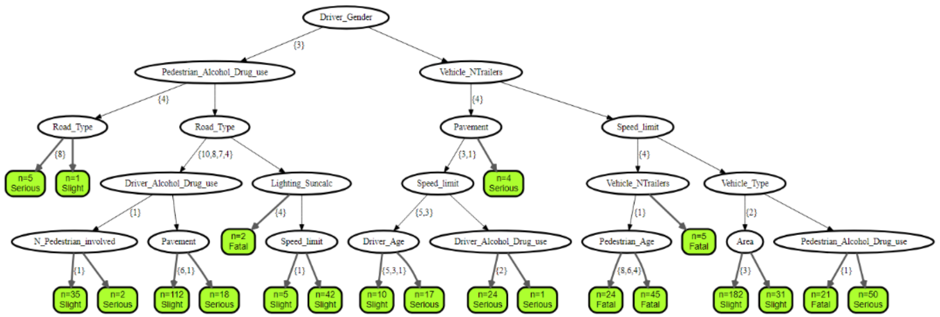
29



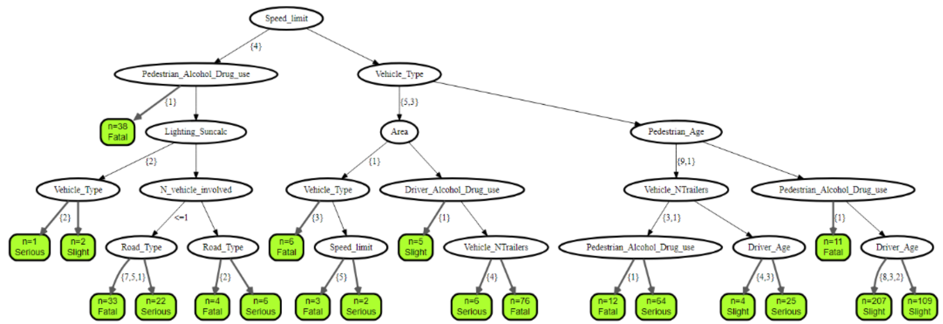
30



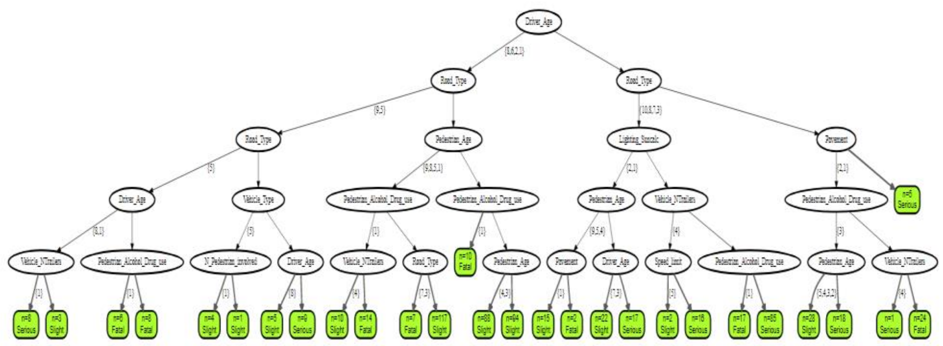
31



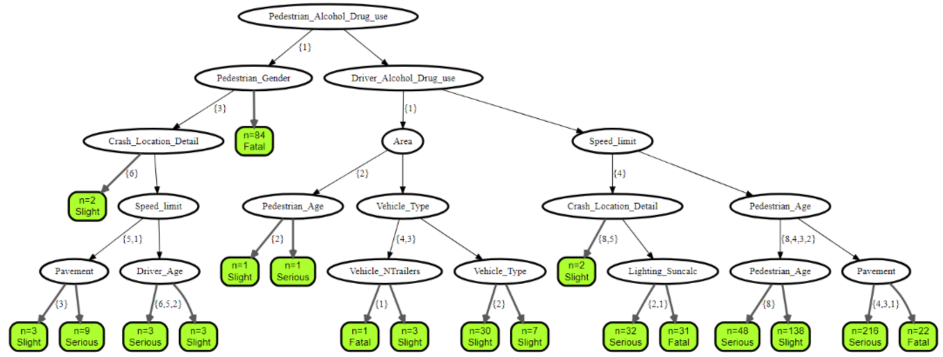
32



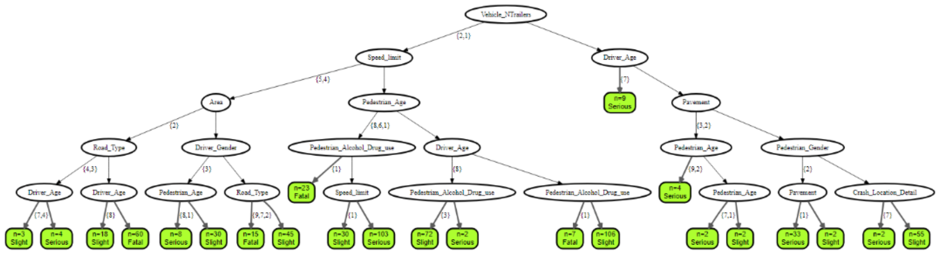
33



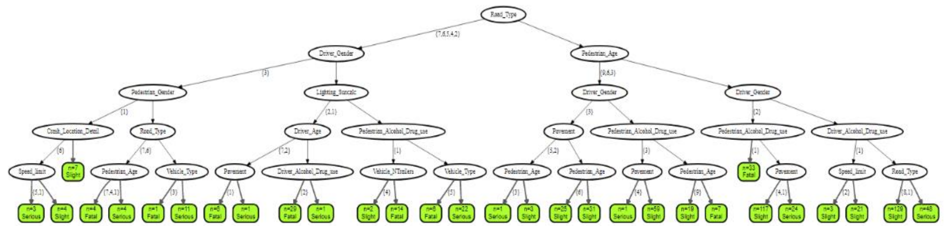
34


Pedestrian_Alcohol_Drug_use
{1}
Pedestrian_Gender
Driver_Alcohol_Drug_use
{3}
Crash_Location_Detail
n=84 Fatal
{1}
Area
Speed_limit
{6}
{2}
{4}
n=2 Slight
Speed_limit
Pedestrian_Age
Vehicle_Type
Crash_Location_Detail
Pedestrian_Age
{5,1}
{2}
{4,3}
{8,5}
{8,4,3,2}
Pavement
Driver_Age
n=1 Slight
n=1 Serious
Vehicle_NTrailers
Vehicle_Type
n=2 Slight
Lighting_Suncalc
Pedestrian_Age
Pavement
{3}
{6,5,2}
{1}
{2}
{2,1}
{8}
{4,3,1}
n=3 Slight
n=9 Serious
n=3 Serious
n=3 Slight
n=1 Fatal
n=3 Slight
n=30 Slight
n=7 Slight
n=32 Serious
n=31 Fatal
n=48 Serious
n=138 Slight
n=216 Serious
n=22 Fatal

35

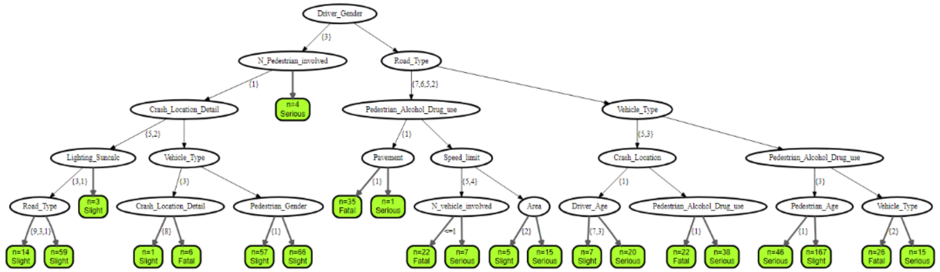


36

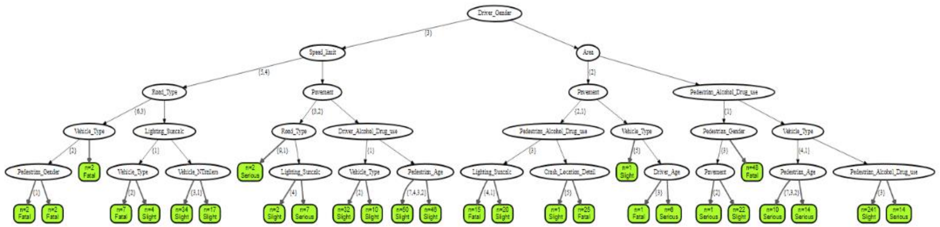


37

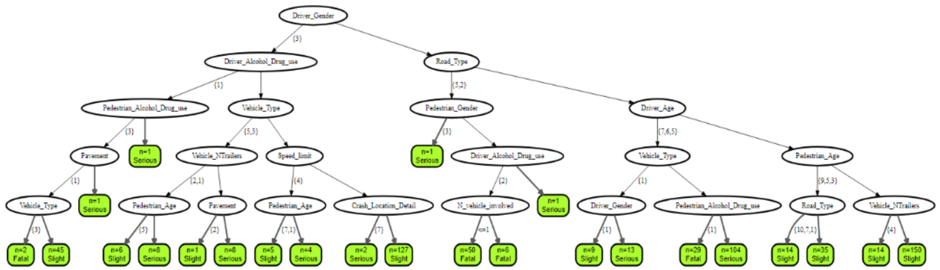
38



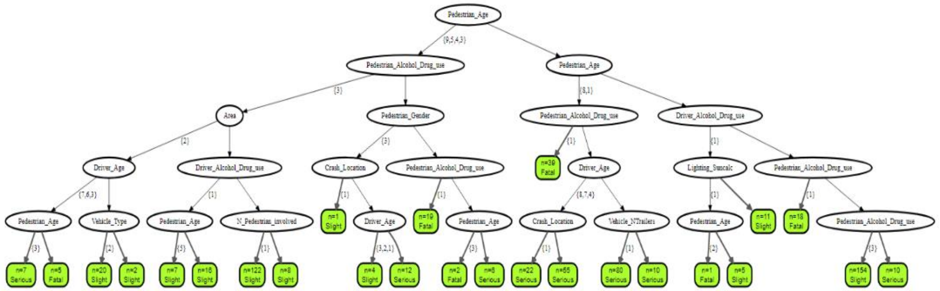
39



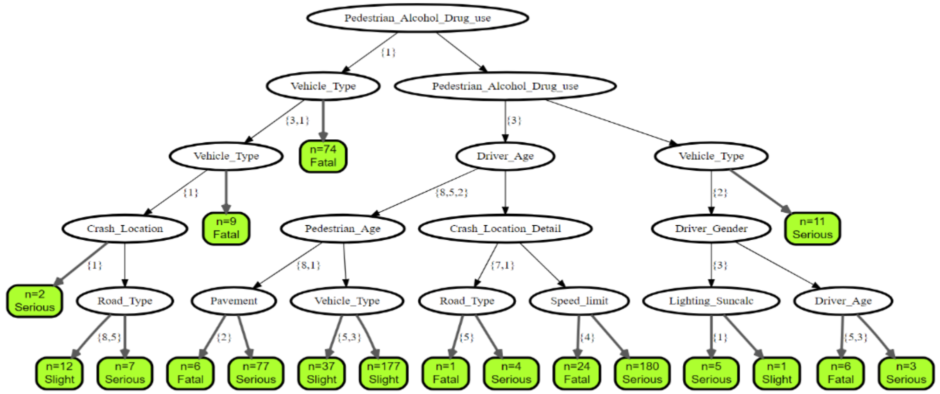
40



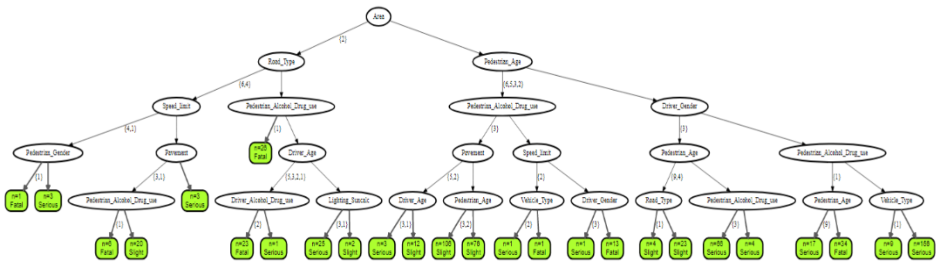
41



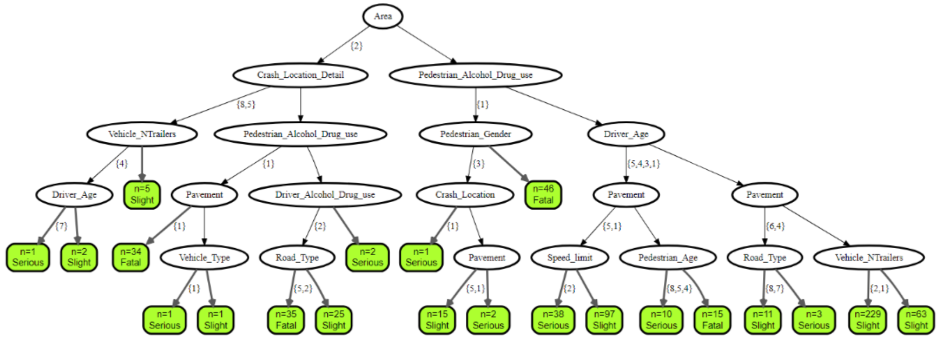
42


Pedestrian_Alcohol_Drug_use
{1}
Vehicle_Type
Pedestrian_Alcohol_Drug_use
{3,1}
{3}
Vehicle_Type
n=74 Fatal
Driver_Age
Vehicle_Type
{1}
n=9 Fatal
{8,5,2}
{2}
n=11 Serious
Crash_Location
Pedestrian_Age
Crash_Location_Detail
Driver_Gender
{1}
{8,1}
{7,1}
{3}
n=2 Serious
Road_Type
Pavement
Vehicle_Type
Road_Type
Speed_limit
Lighting_Suncalc
Driver_Age
{8,5}
{2}
{5,3}
{5}
{4}
{1}
{5,3}
n=12 Slight
n=7 Serious
n=6 Fatal
n=77 Serious
n=37 Slight
n=177 Slight
n=1 Fatal
n=4 Serious
n=24 Fatal
n=180 Serious
n=5 Serious
n=1 Slight
n=6 Fatal
n=3 Serious

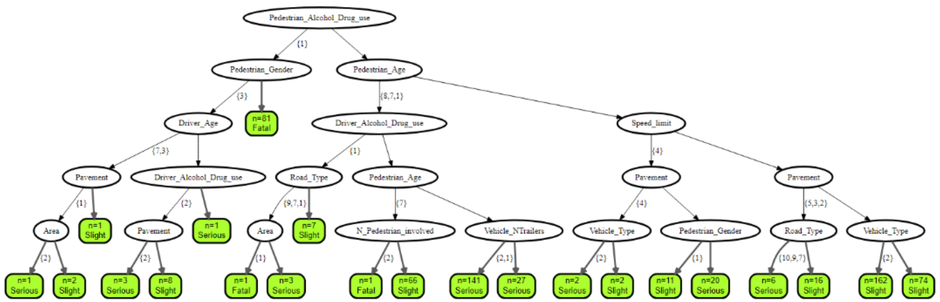
43



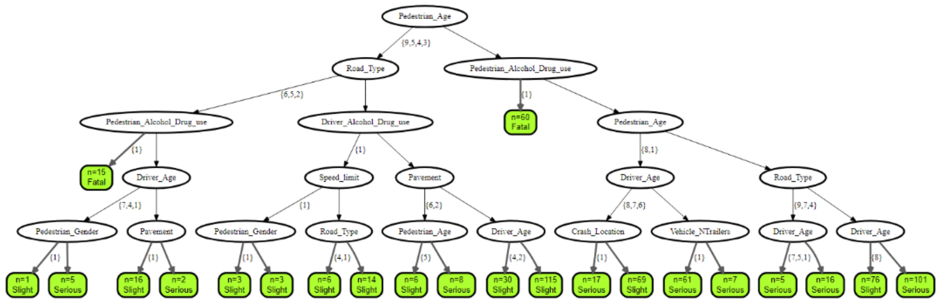
44


Area
{2}
Crash_Location_Detail
Pedestrian_Alcohol_Drug_use
{8,5}
{1}
Vehicle_NTrailers
Pedestrian_Alcohol_Drug_use
Pedestrian_Gender
Driver_Age
{4}
{1}
{3}
{5,4,3,1}
Driver_Age
n=5 Slight
Pavement
Driver_Alcohol_Drug_use
Crash_Location
n=46 Fatal
Pavement
Pavement
{7}
{1}
{2}
{1}
{5,1}
{6,4}
n=1 Serious
n=2 Slight
n=34 Fatal
Vehicle_Type
Road_Type
n=2 Serious
n=1 Serious
Pavement
Speed_limit
Pedestrian_Age
Road_Type
Vehicle_NTrailers
{1}
{5,2}
{5,1}
{2}
{8,5,4}
{8,7}
{2,1}
n=1 Serious
n=1 Slight
n=35 Fatal
n=25 Slight
n=15 Slight
n=2 Serious
n=38 Serious
n=97 Slight
n=10 Serious
n=15 Fatal
n=11 Slight
n=3 Serious
n=229 Slight
n=63 Slight

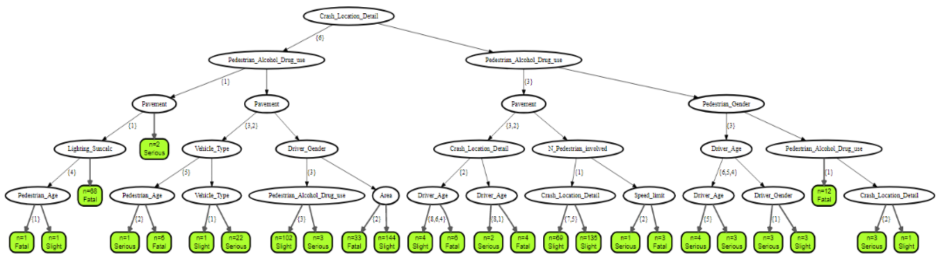
45



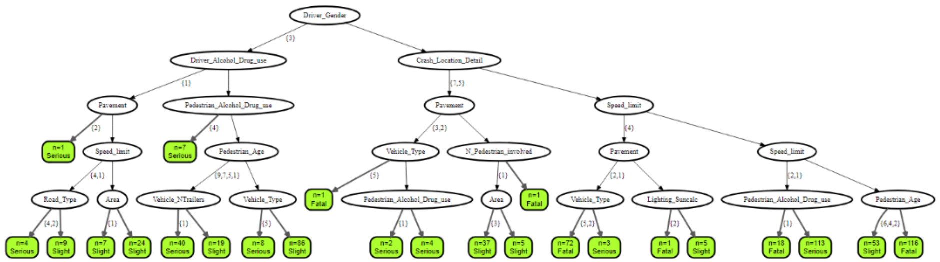
46



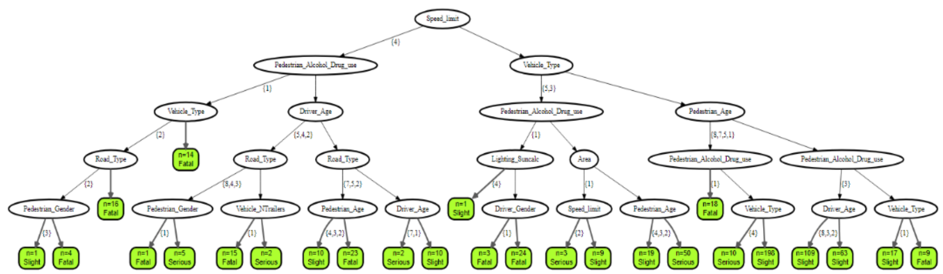
47



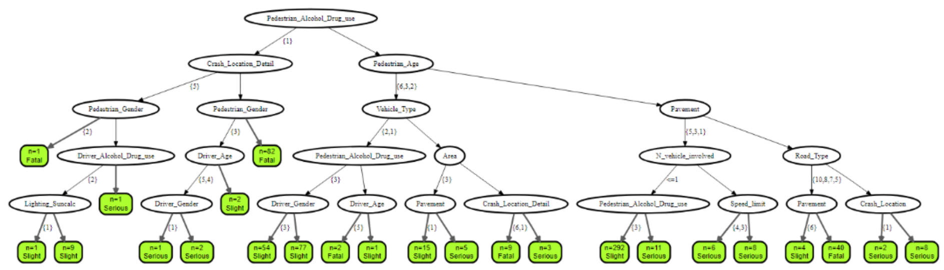
48



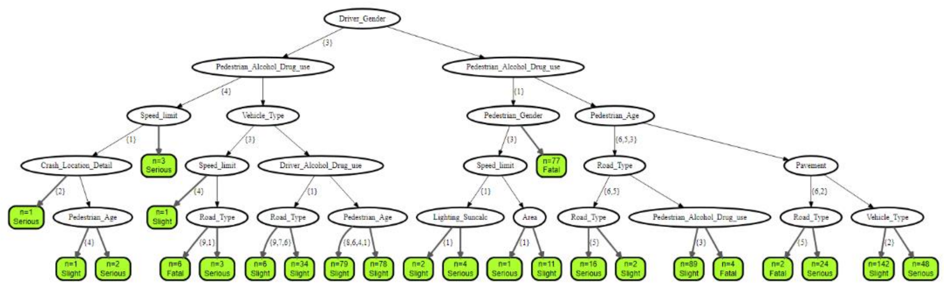
49



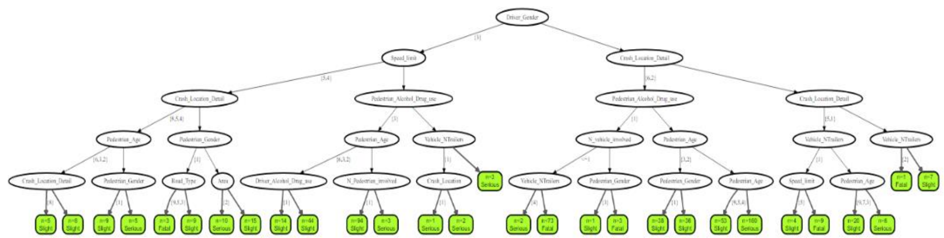
50



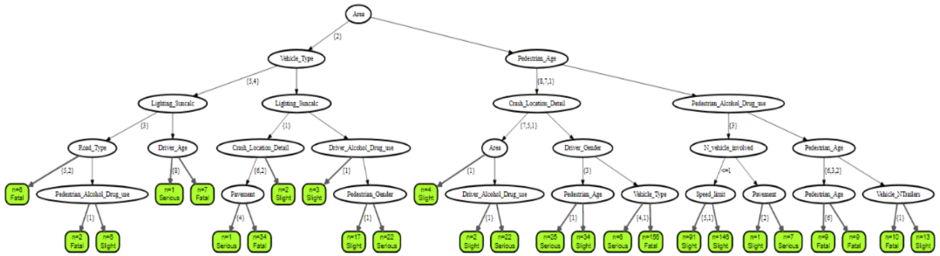
51



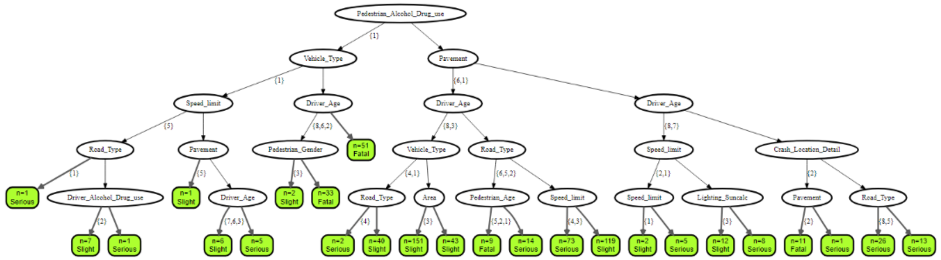
52



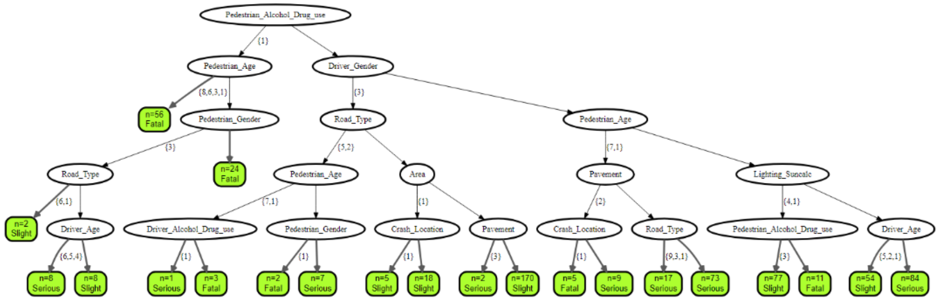
53



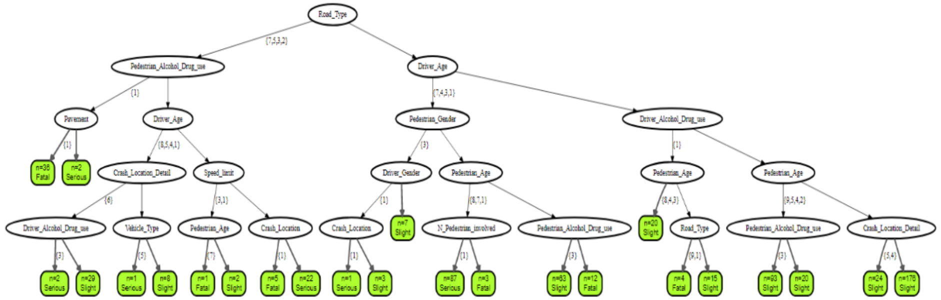
54



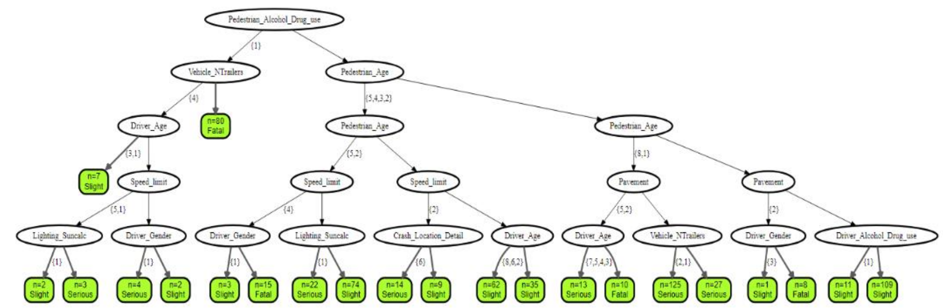
55



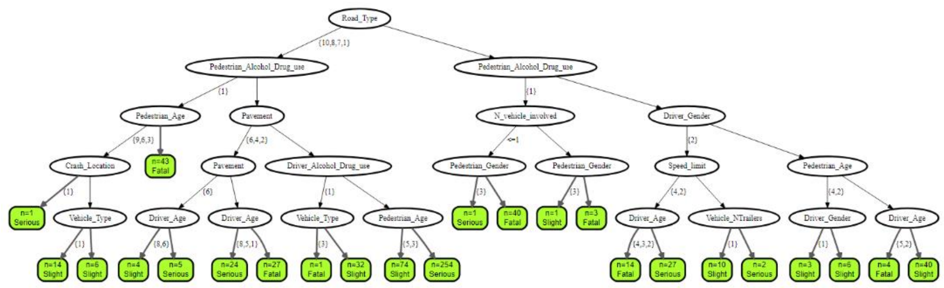
56



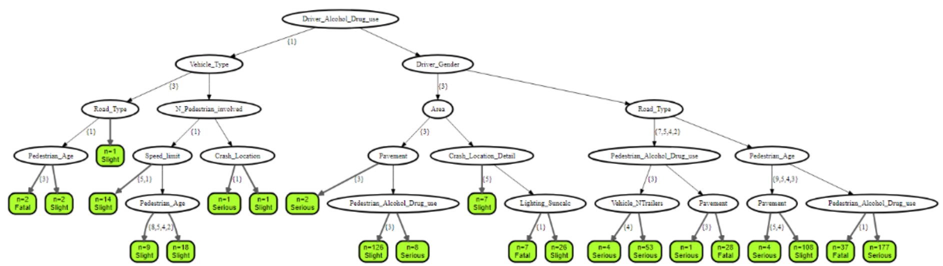
57



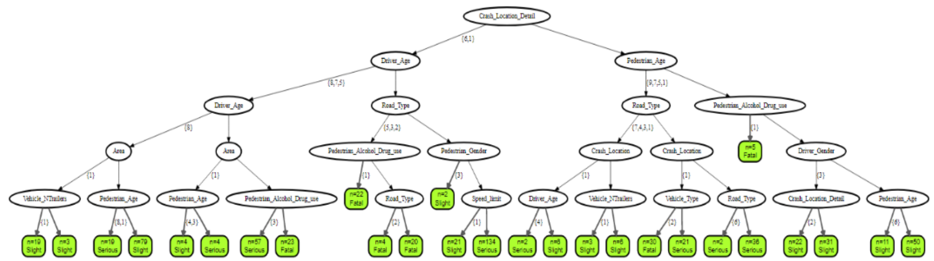
58



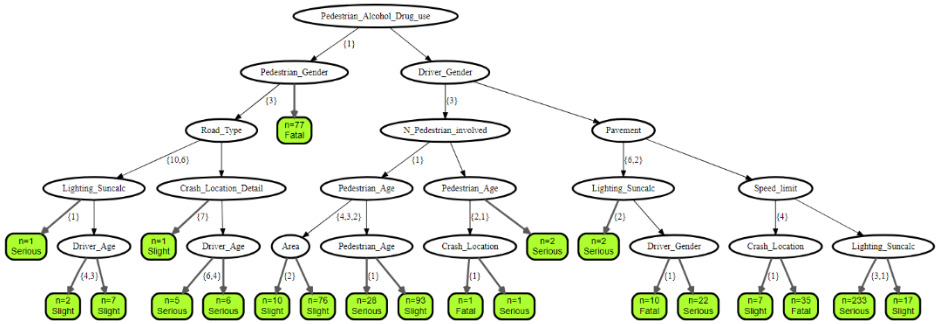
59



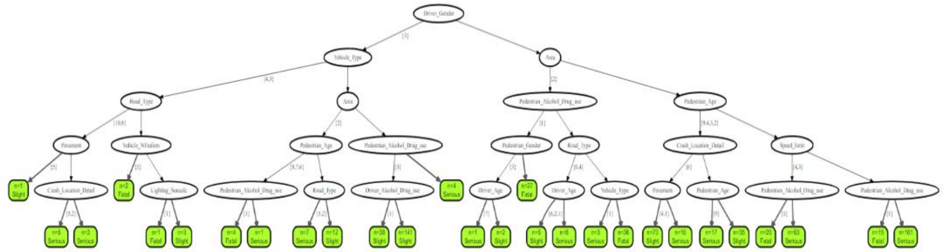
60



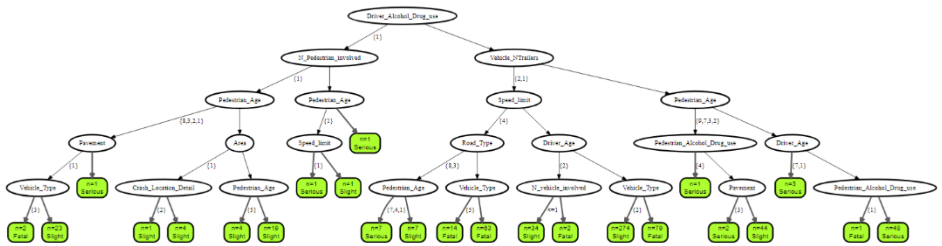
61


Pedestrian_Alcohol_Drug_use
{1}
Pedestrian_Gender
Driver_Gender
{3}
{3}
Road_Type
n=77 Fatal
N_Pedestrian_involved
Pavement
{10,6}
{1}
{6,2}
Lighting_Suncalc
Crash_Location_Detail
Pedestrian_Age
Pedestrian_Age
Lighting_Suncalc
Speed_limit
{1}
{7}
{4,3,2}
{2,1}
{2}
{4}
n=1 Serious
Driver_Age
n=1 Slight
Driver_Age
Area
Pedestrian_Age
Crash_Location
n=2 Serious
n=2 Serious
Driver_Gender
Crash_Location
Lighting_Suncalc
{4,3}
{6,4}
{2}
{1}
{1}
{1}
{1}
{3,1}
n=2 Slight
n=7 Slight
n=5 Serious
n=6 Serious
n=10 Slight
n=76 Slight
n=28 Serious
n=93 Slight
n=1 Fatal
n=1 Serious
n=10 Fatal
n=22 Serious
n=7 Slight
n=35 Fatal
n=233 Serious
n=17 Slight

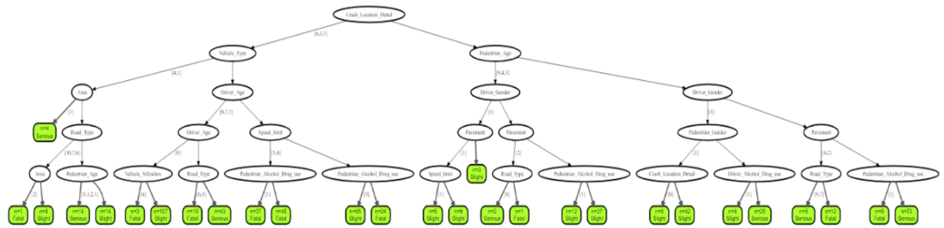
62



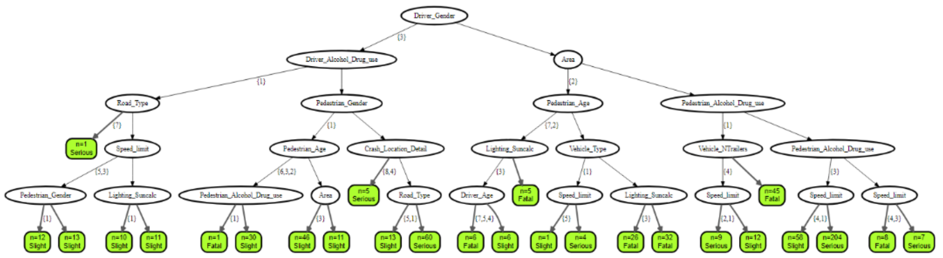
63



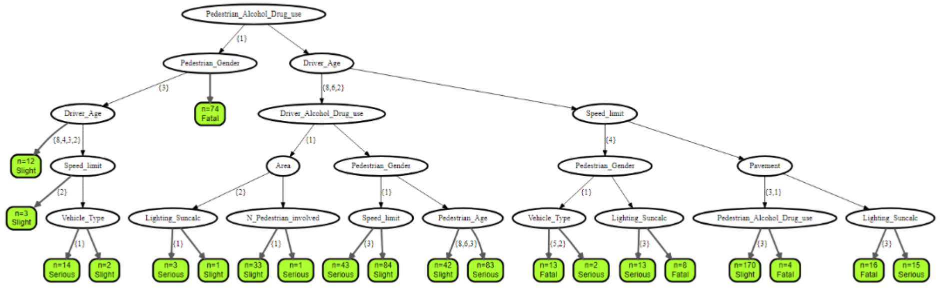
64



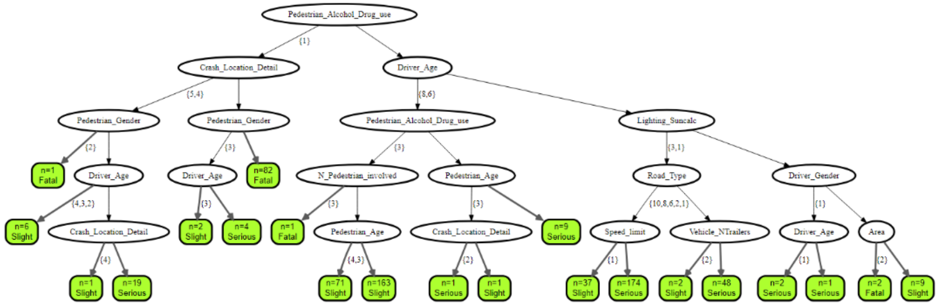
65



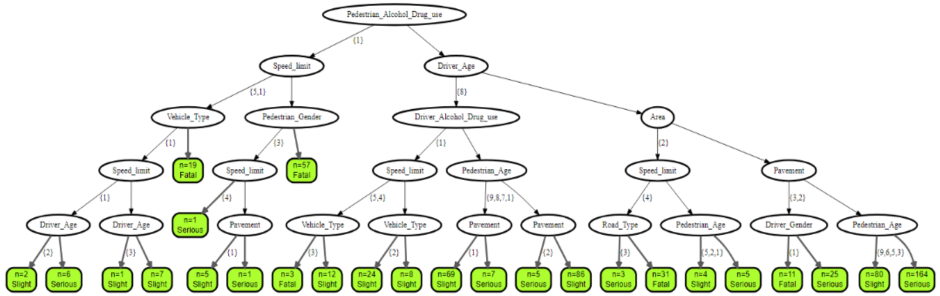
66



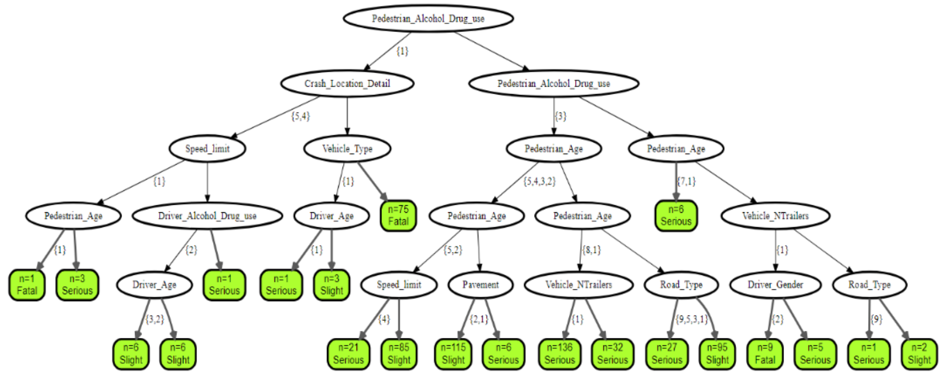
67



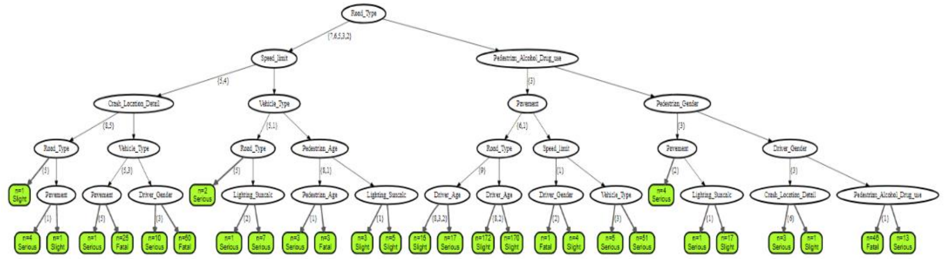
68



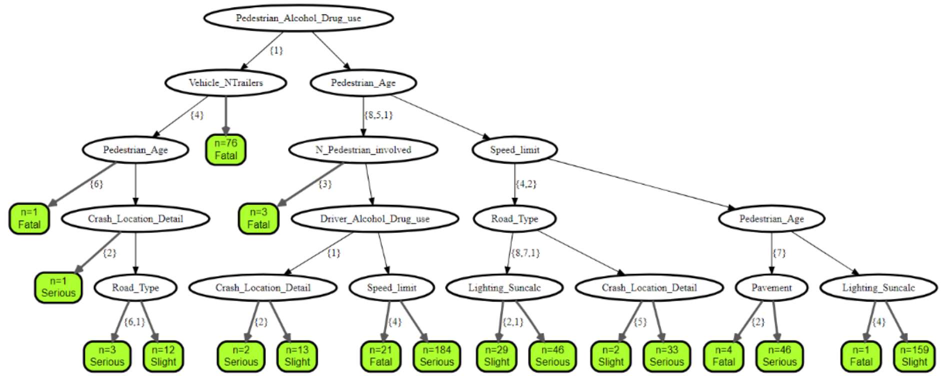
69


Pedestrian_Alcohol_Drug_use
{1}
Crash_Location_Detail
Pedestrian_Alcohol_Drug_use
{5,4}
{3}
Speed_limit
Vehicle_Type
Pedestrian_Age
Pedestrian_Age
{1}
{1}
{5,4,3,2}
{7,1}
Pedestrian_Age
Driver_Alcohol_Drug_use
Driver_Age
n=75 Fatal
Pedestrian_Age
Pedestrian_Age
n=6 Serious
Vehicle_NTrailers
{1}
{2}
{1}
{5,2}
{8,1}
{1}
n=1 Fatal
n=3 Serious
Driver_Age
n=1 Serious
n=1 Serious
n=3 Slight
Speed_limit
Pavement
Vehicle_NTrailers
Road_Type
Driver_Gender
Road_Type
{3,2}
{4}
{2,1}
{1}
{9,5,3,1}
{2}
{9}
n=6 Slight
n=6 Slight
n=21 Serious
n=85 Slight
n=115 Slight
n=6 Serious
n=136 Serious
n=32 Serious
n=27 Serious
n=95 Slight
n=9 Fatal
n=5 Serious
n=1 Serious
n=2 Slight

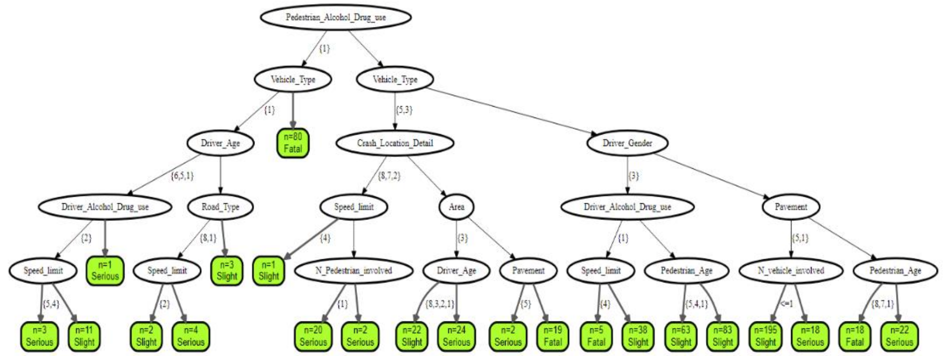
70



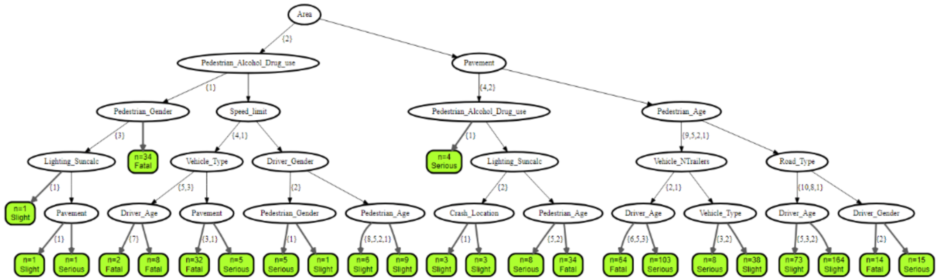
71


Pedestrian_Alcohol_Drug_use
{1}
Vehicle_NTrailers
Pedestrian_Age
{4}
{8,5,1}
Pedestrian_Age
n=76 Fatal
N_Pedestrian_involved
Speed_limit
{6}
{3}
{4,2}
n=1 Fatal
Crash_Location_Detail
n=3 Fatal
Driver_Alcohol_Drug_use
Road_Type
Pedestrian_Age
{2}
{1}
{8,7,1}
{7}
n=1 Serious
Road_Type
Crash_Location_Detail
Speed_limit
Lighting_Suncalc
Crash_Location_Detail
Pavement
Lighting_Suncalc
{6,1}
{2}
{4}
{2,1}
{5}
{2}
{4}
n=3 Serious
n=12 Slight
n=2 Serious
n=13 Slight
n=21 Fatal
n=184 Serious
n=29 Slight
n=46 Serious
n=2 Slight
n=33 Serious
n=4 Fatal
n=46 Serious
n=1 Fatal
n=159 Slight

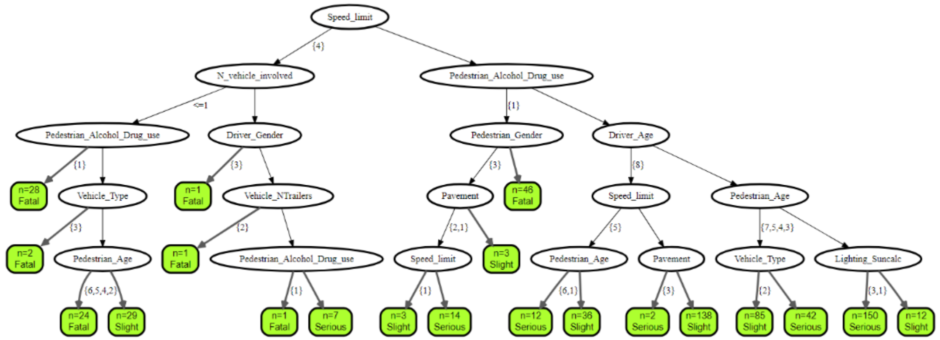
72


Pedestrian_Alcohol_Drug_use
{1}
Vehicle_Type
Vehicle_Type
{1}
{5,3}
Driver_Age
n=80 Fatal
Crash_Location_Detail
Driver_Gender
{6,5,1}
{8,7,2}
{3}
Driver_Alcohol_Drug_use
Road_Type
Speed_limit
Area
Driver_Alcohol_Drug_use
Pavement
{2}
{8,1}
{4}
{3}
{1}
{5,1}
Speed_limit
n=1 Serious
Speed_limit
n=3 Slight
n=1 Slight
N_Pedestrian_involved
Driver_Age
Pavement
Speed_limit
Pedestrian_Age
N_vehicle_involved
Pedestrian_Age
{5,4}
{2}
{1}
{8,3,2,1}
{5}
{4}
{5,4,1}
<=1
{8,7,1}
n=3 Serious
n=11 Slight
n=2 Slight
n=4 Serious
n=20 Serious
n=2 Serious
n=22 Slight
n=24 Serious
n=2 Serious
n=19 Fatal
n=5 Fatal
n=38 Slight
n=63 Slight
n=83 Slight
n=195 Slight
n=18 Serious
n=18 Fatal
n=22 Serious

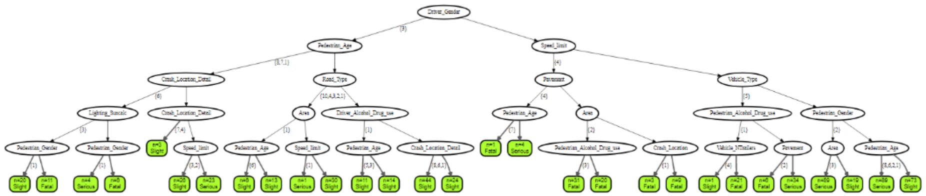
73



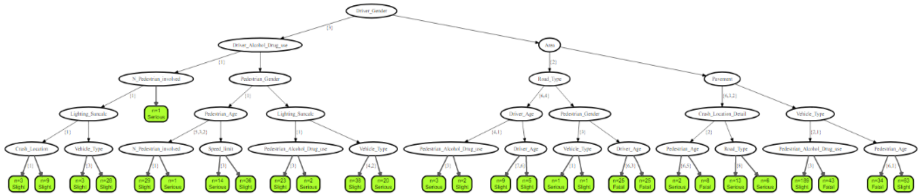
74


Speed_limit
{4}
N_vehicle_involved
Pedestrian_Alcohol_Drug_use
<=1
Pedestrian_Alcohol_Drug_use
Driver_Gender
{1}
Pedestrian_Gender
Driver_Age
{1}
n=28 Fatal
Vehicle_Type
{3}
n=1 Fatal
Vehicle_NTrailers
{3}
n=46 Fatal
{8}
Pavement
Speed_limit
Pedestrian_Age
{3}
n=2 Fatal
Pedestrian_Age
{2}
n=1 Fatal
Pedestrian_Alcohol_Drug_use
{2,1}
Speed_limit
n=3 Slight
{5}
Pedestrian_Age
Pavement
{7,5,4,3}
Vehicle_Type
Lighting_Suncalc
{6,5,4,2}
n=24 Fatal
n=29 Slight
{1}
n=1 Fatal
n=7 Serious
{1}
n=3 Slight
n=14 Serious
{6,1}
n=12 Serious
n=36 Slight
{3}
n=2 Serious
n=138 Slight
{2}
n=85 Slight
n=42 Serious
{3,1}
n=150 Serious
n=12 Slight

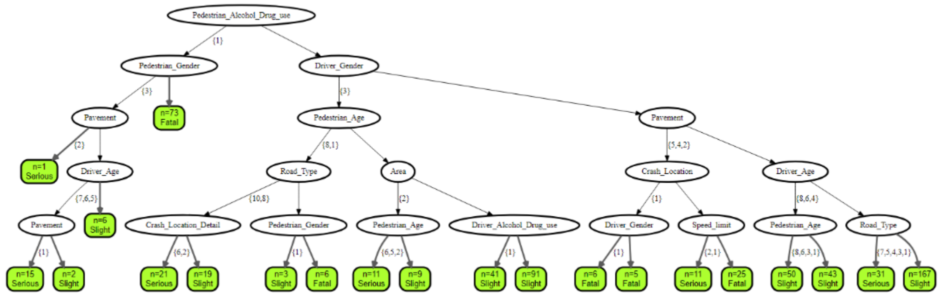
75



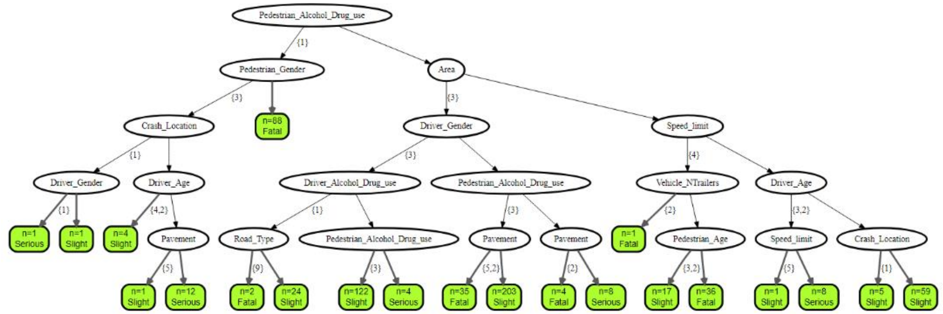
76



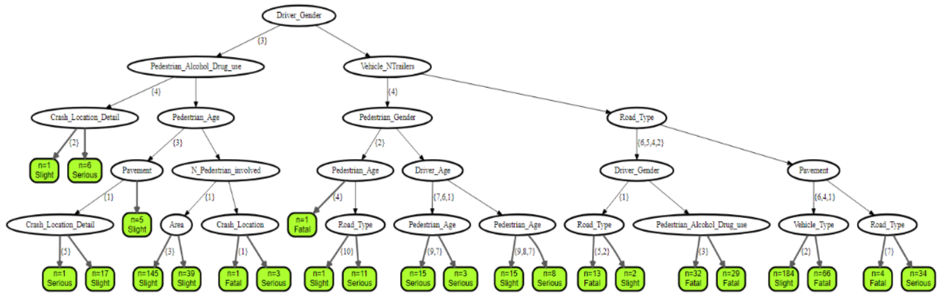
77



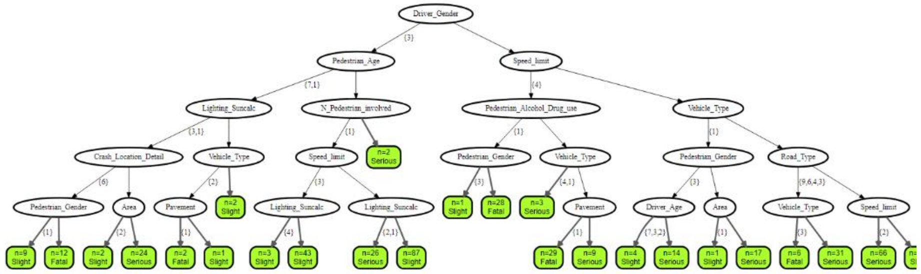
78



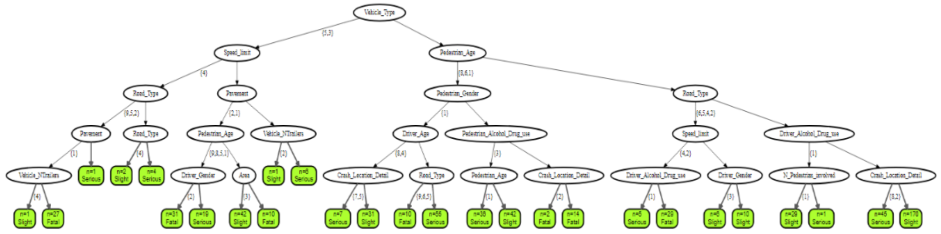
79



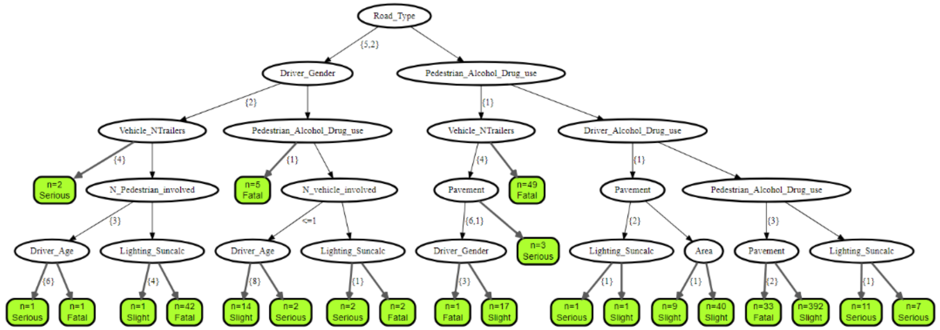
80



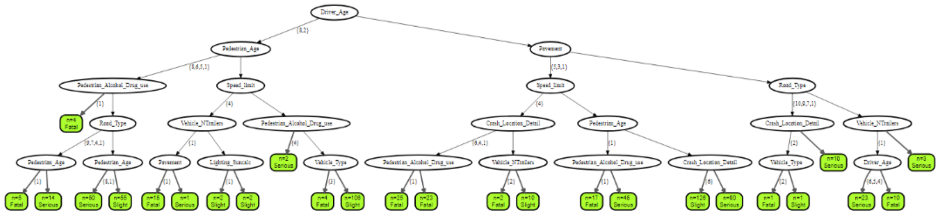
81



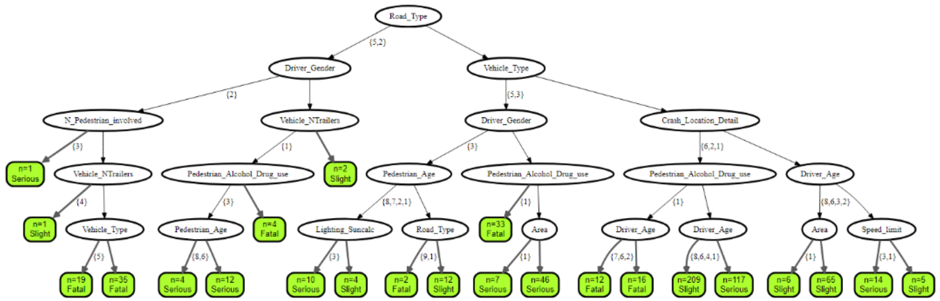
82


Road_Type
{5,2}
Driver_Gender
Pedestrian_Alcohol_Drug_use
{2}
{1}
Vehicle_NTrailers
Pedestrian_Alcohol_Drug_use
Vehicle_NTrailers
Driver_Alcohol_Drug_use
{4}
{1}
{4}
{1}
n=2 Serious
N_Pedestrian_involved
n=5 Fatal
N_vehicle_involved
Pavement
n=49 Fatal
Pavement
Pedestrian_Alcohol_Drug_use
{3}
<=1
{6,1}
{2}
{3}
Driver_Age
Lighting_Suncalc
Driver_Age
Lighting_Suncalc
Driver_Gender
n=3 Serious
Lighting_Suncalc
Area
Pavement
Lighting_Suncalc
{6}
{4}
{8}
{1}
{3}
{1}
{1}
{2}
{1}
n=1 Serious
n=1 Fatal
n=1 Slight
n=42 Fatal
n=14 Slight
n=2 Serious
n=2 Serious
n=2 Fatal
n=1 Fatal
n=17 Slight
n=1 Serious
n=1 Slight
n=9 Slight
n=40 Slight
n=33 Fatal
n=392 Slight
n=11 Serious
n=7 Serious

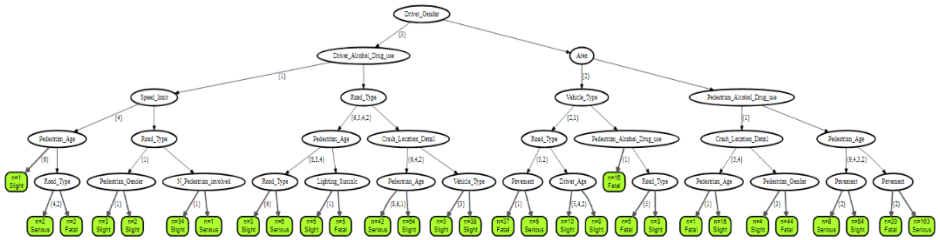
83



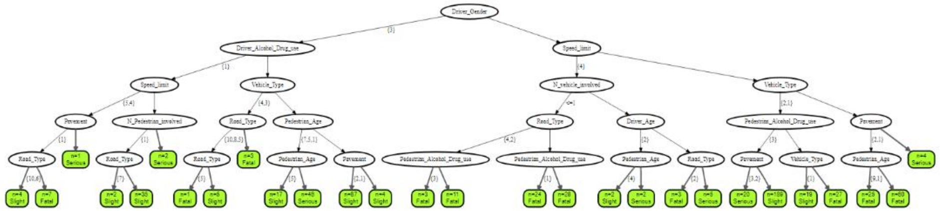
84


Road_Type
{5,2}
Driver_Gender
Vehicle_Type
{2}
{5,3}
N_Pedestrian_involved
Vehicle_NTrailers
Driver_Gender
Crash_Location_Detail
{3}
{1}
{3}
{6,2,1}
n=1 Serious
Vehicle_NTrailers
Pedestrian_Alcohol_Drug_use
n=2 Slight
Pedestrian_Age
Pedestrian_Alcohol_Drug_use
Pedestrian_Alcohol_Drug_use
Driver_Age
{4}
{3}
{8,7,2,1}
{1}
{1}
{8,6,3,2}
n=1 Slight
Vehicle_Type
Pedestrian_Age
n=4 Fatal
Lighting_Suncalc
Road_Type
n=33 Fatal
Area
Driver_Age
Driver_Age
Area
Speed_limit
{5}
{8,6}
{3}
{9,1}
{1}
{7,6,2}
{8,6,4,1}
{1}
{3,1}
n=19 Fatal
n=35 Fatal
n=4 Serious
n=12 Serious
n=10 Serious
n=4 Slight
n=2 Fatal
n=12 Slight
n=7 Serious
n=46 Serious
n=12 Fatal
n=16 Fatal
n=209 Slight
n=117 Serious
n=6 Slight
n=65 Slight
n=14 Serious
n=5 Slight

85

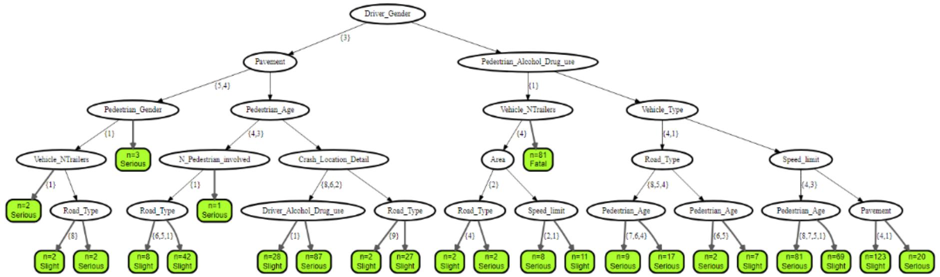


86

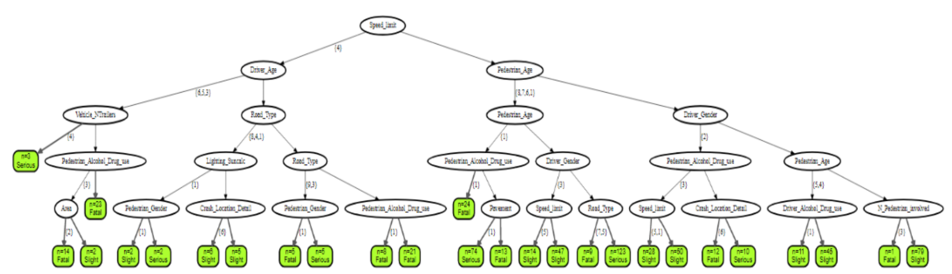


87

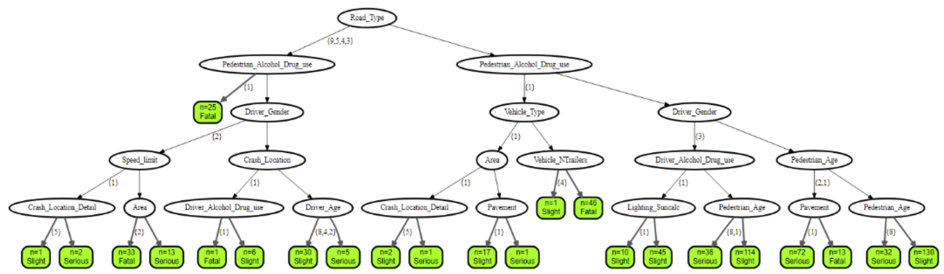
88



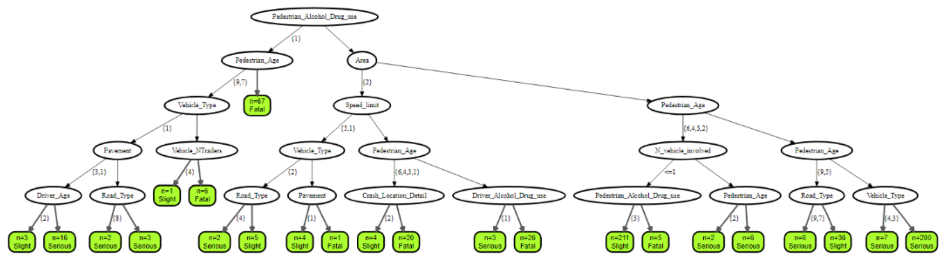
89



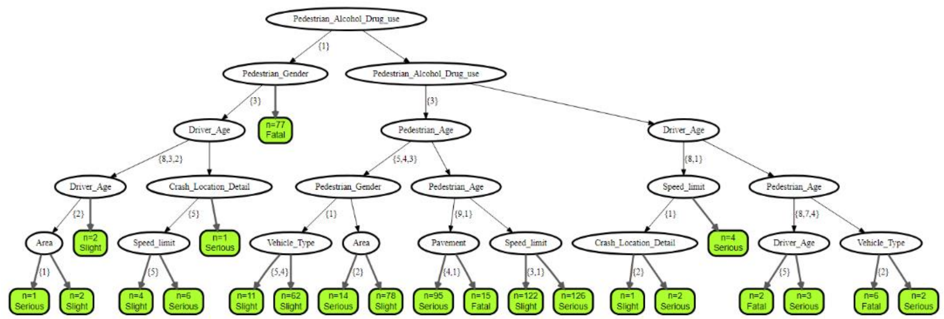
90



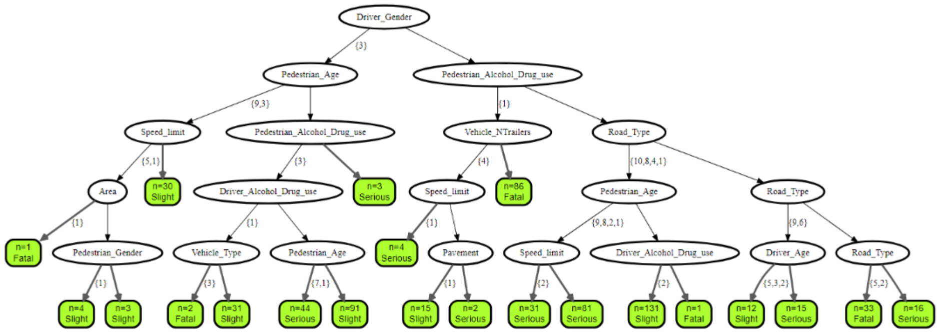
91



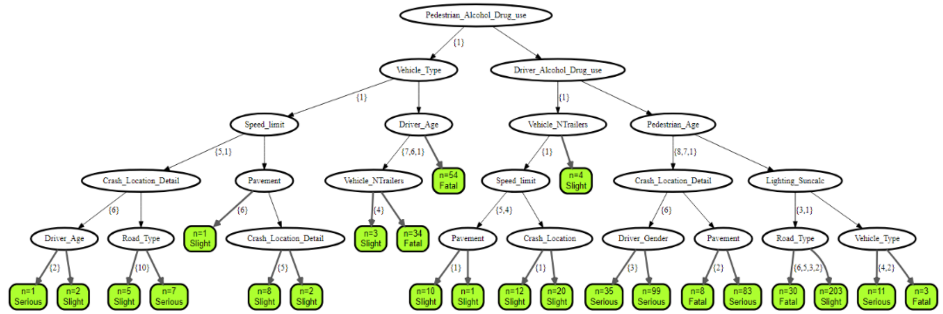
92


Pedestrian_Alcohol_Drug_use
{1}
Pedestrian_Gender
Pedestrian_Alcohol_Drug_use
{3}
n=77 Fatal
{3}
Driver_Age
Pedestrian_Age
Driver_Age
{8,3,2}
{5,4,3}
{8,1}
Driver_Age
Crash_Location_Detail
Pedestrian_Gender
Pedestrian_Age
Speed_limit
Pedestrian_Age
{2}
{5}
{1}
{9,1}
{1}
{8,7,4}
Area
n=2 Slight
Speed_limit
n=1 Serious
Vehicle_Type
Area
Pavement
Speed_limit
Crash_Location_Detail
n=4 Serious
Driver_Age
Vehicle_Type
{1}
{5}
{5,4}
{2}
{4,1}
{3,1}
{2}
{5}
{2}
n=1 Serious
n=2 Slight
n=4 Slight
n=6 Serious
n=11 Slight
n=62 Slight
n=14 Serious
n=78 Slight
n=95 Serious
n=15 Fatal
n=122 Slight
n=126 Serious
n=1 Slight
n=2 Serious
n=2 Fatal
n=3 Serious
n=6 Fatal
n=2 Serious

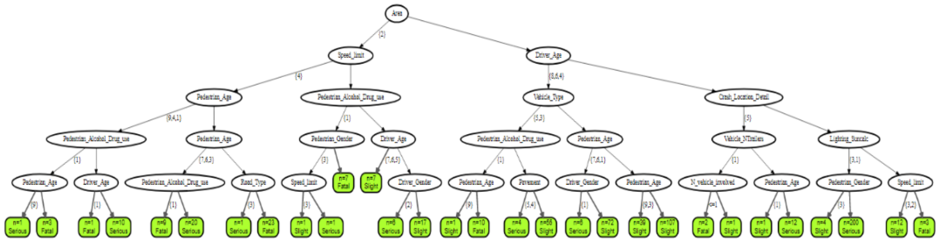
93


Driver_Gender
Pedestrian_Age
Pedestrian_Alcohol_Drug_use
Speed_limit
Pedestrian_Alcohol_Drug_use
Vehicle_NTrailers
Road_Type
Area
n=30 Slight
Driver_Alcohol_Drug_use
n=3 Serious
Speed_limit
n=86 Fatal
Pedestrian_Age
Road_Type
n=1 Fatal
Pedestrian_Gender
Vehicle_Type
Pedestrian_Age
n=4 Serious
Pavement
Speed_limit
Driver_Alcohol_Drug_use
Driver_Age
Road_Type
n=4 Slight
n=3 Slight
n=2 Fatal
n=31 Slight
n=44 Serious
n=91 Slight
n=15 Slight
n=2 Serious
n=31 Serious
n=81 Serious
n=131 Slight
n=1 Fatal
n=12 Slight
n=15 Serious
n=33 Fatal
n=16 Serious

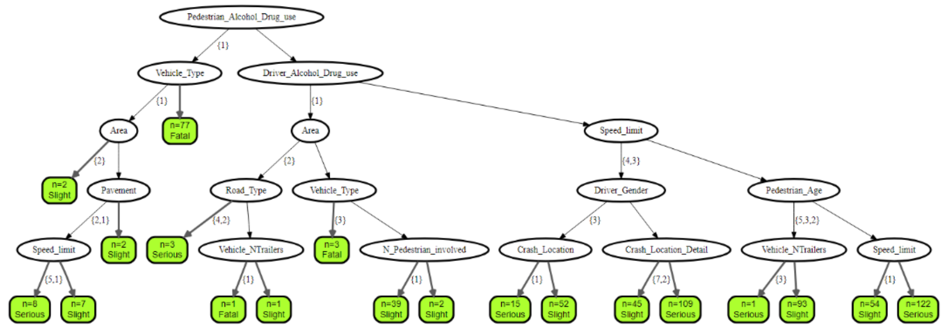
94


Pedestrian_Alcohol_Drug_use
Vehicle_Type
Driver_Alcohol_Drug_use
Speed_limit
Driver_Age
Vehicle_NTrailers
Pedestrian_Age
Crash_Location_Detail
Pavement
Vehicle_NTrailers
n=54 Fatal
Speed_limit
n=4 Slight
Crash_Location_Detail
Lighting_Suncalc
Driver_Age
Road_Type
n=1 Slight
Crash_Location_Detail
n=3 Slight
n=34 Fatal
Pavement
Crash_Location
Driver_Gender
Pavement
Road_Type
Vehicle_Type

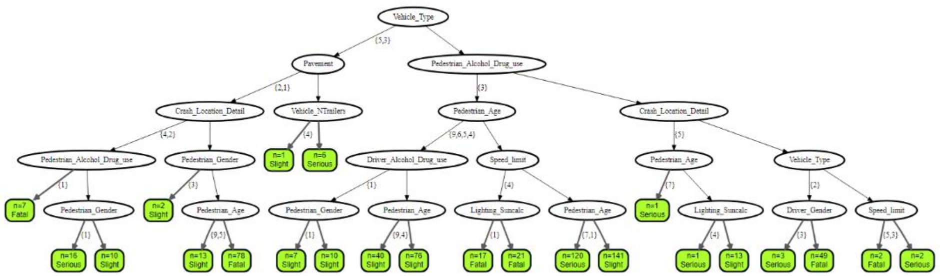
95



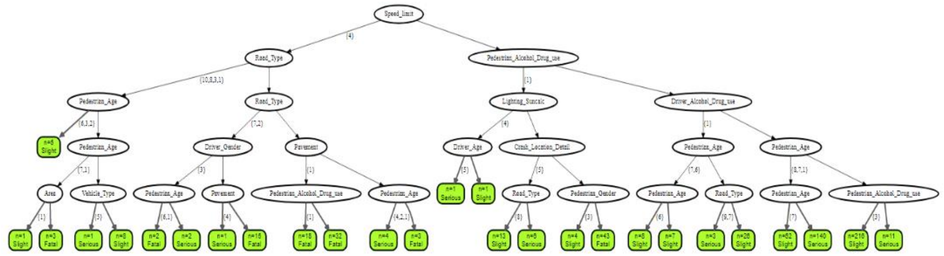
96


Pedestrian_Alcohol_Drug_use
Vehicle_Type
Driver_Alcohol_Drug_use
Area
n=77 Fatal
Area
Speed_limit
n=2 Slight
Pavement
Road_Type
Vehicle_Type
Driver_Gender
Pedestrian_Age
Speed_limit
n=2 Slight
n=3 Serious
Vehicle_NTrailers
n=3 Fatal
N_Pedestrian_involved
Crash_Location
Crash_Location_Detail
Vehicle_NTrailers
Speed_limit
n=8 Serious
n=7 Slight
n=1 Fatal
n=1 Slight
n=39 Slight
n=2 Slight
n=15 Serious
n=52 Slight
n=45 Slight
n=109 Serious
n=1 Serious
n=93 Slight
n=54 Slight
n=122 Serious

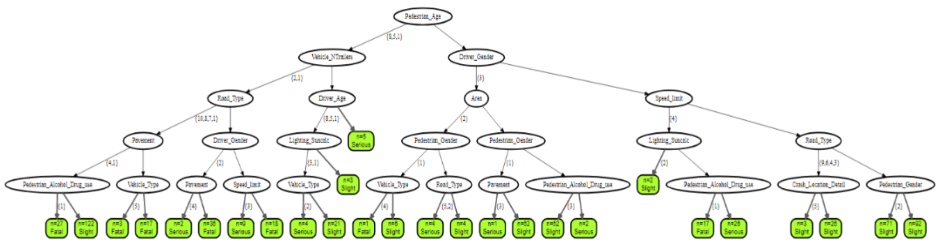
97



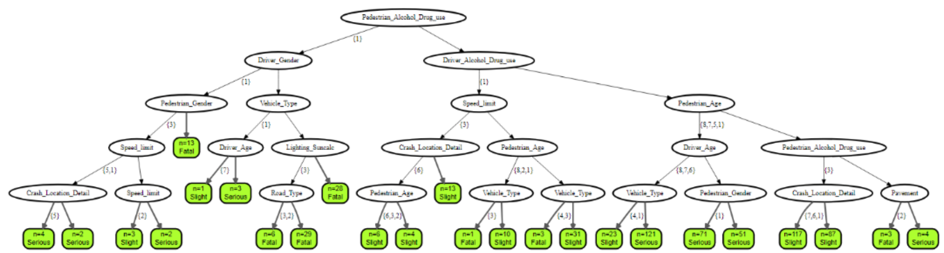
98



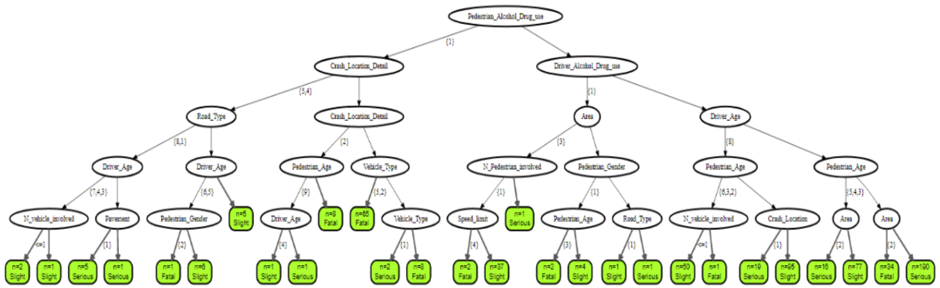
99



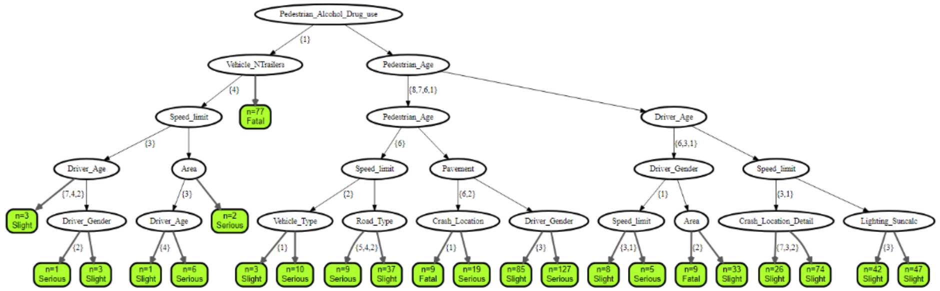
100



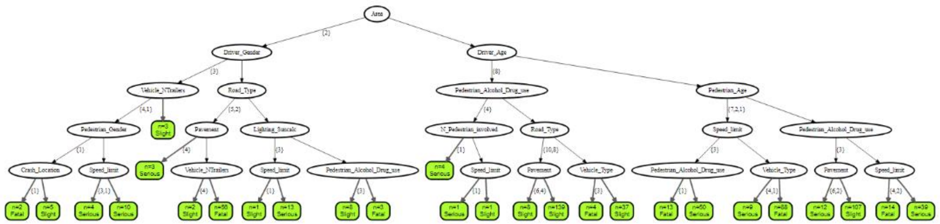
101



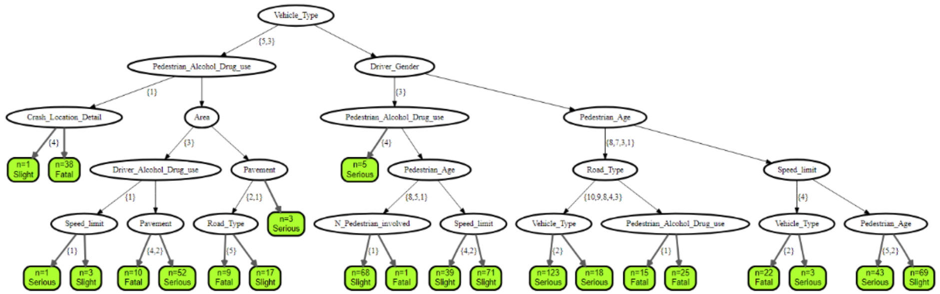
102

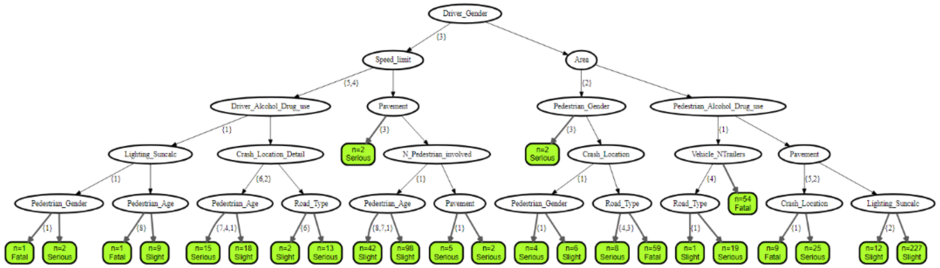


103



104



105	
Driver_Gender {3} Speed_limit Area {5,4} {2} Driver_Alcohol_Drug_use Pavement Pedestrian_Gender Pedestrian_Alcohol_Drug_use {1} {3} {3} {1} Lighting_Suncalc Crash_Location_Detail n=2 Serious N_Pedestrian_involved n=2 Serious Crash_Location Vehicle_NTrailers Pavement {1} {6,2} {1} {1} {4} {5,2} Pedestrian_Gender Pedestrian_Age Pedestrian_Age Road_Type Pedestrian_Age Pavement Pedestrian_Gender Road_Type Road_Type n=54 Fatal Crash_Location Lighting_Suncalc {1} {8} {7,4,1} {6} {8,7,1} {1} {1} {4,3} {1} {1} {2} n=1 Fatal n=2 Serious n=1 Fatal n=9 Slight n=15 Serious n=18 Slight n=2 Slight n=13 Serious n=42 Slight n=98 Slight n=5 Serious n=2 Serious n=4 Serious n=6 Slight n=8 Serious n=59 Fatal n=1 Slight n=19 Serious n=9 Fatal n=25 Serious n=12 Slight n=227 Slight	

Association rules

Rules with fatal as consequent

Table 158 – Association rules with roadway characteristics (Road type) as first antecedent and fatal crashes as consequent, Sweden.

ID Rule	Rules with roadway characteristics (Road type) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
1	Road Type=Motorway	0.21	24.39	10.84	n.a.
2	Road Type=Motorway & Driver Alcohol/Drug use=No	0.21	26.32	11.70	1.08
3	Road Type=Motorway & Driver Alcohol/Drug use=No & N. Pedestrian involved=1	0.21	28.17	12.52	1.07
4	Road Type=Motorway & Driver Alcohol/Drug use=No & N. Pedestrian involved=1 & Crash Location Detail=Road section	0.20	29.69	13.20	1.05
5	Road Type=Motorway & N. Pedestrian involved=1	0.21	25.97	11.55	1.06
6	Road Type=Motorway & Pavement=Dry	0.20	25.68	11.42	1.05
7	Road Type=Rural National	0.48	12.82	5.70	n.a.
8	Road Type=Rural National & Lighting=Darkness	0.31	20.57	9.14	1.60
9	Road Type=Rural National & Lighting=Darkness & Driver Gender=Male	0.24	28.75	12.78	1.40
10	Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Day of Week=Weekday	0.20	42.22	18.77	1.47
11	Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Day of Week=Weekday & Speed Limit≥60	0.20	46.34	20.60	1.64
12	Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Pavement=Dry	0.23	31.43	13.97	1.09
13	Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Pavement=Dry & Speed Limit≥60	0.22	33.87	15.06	1.08
14	Road Type=Rural National & Lighting=Darkness & Driver Gender=Male & Speed Limit≥60	0.23	30.99	13.78	1.08
15	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday	0.25	26.09	11.60	1.27
16	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Pavement=Dry	0.24	30.67	13.64	1.18
17	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Pavement=Dry & Speed Limit≥60	0.24	33.33	14.82	1.09
18	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Speed Limit≥60	0.25	28.24	12.55	1.08
19	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Speed Limit≥60 & Driver Alcohol/Drug use=No	0.25	30.00	13.34	1.06
20	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Crash Location Detail=Road section	0.23	27.85	12.38	1.07
21	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Crash Location Detail=Road section & Speed Limit≥60	0.23	30.56	13.59	1.10
22	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.25	27.59	12.27	1.06
23	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Crash Location=Not at intersection	0.23	27.50	12.23	1.05
24	Road Type=Rural National & Lighting=Darkness & Day of Week=Weekday & Crash Location=Not at intersection & Speed Limit≥60	0.23	30.14	13.40	1.10
25	Road Type=Rural National & Lighting=Darkness & Pedestrian Gender=Male	0.24	25.27	11.24	1.23
26	Road Type=Rural National & Lighting=Darkness & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.24	27.38	12.17	1.08
27	Road Type=Rural National & Lighting=Darkness & Pedestrian Gender=Male & Driver Alcohol/Drug use=No & Speed Limit≥60	0.23	29.73	13.22	1.09
28	Road Type=Rural National & Lighting=Darkness & Pedestrian Gender=Male & Speed Limit≥60	0.23	27.16	12.08	1.07
29	Road Type=Rural National & Lighting=Darkness & Pedestrian Gender=Male & Speed Limit≥60 & Pavement=Dry	0.22	28.77	12.79	1.06
30	Road Type=Rural National & Lighting=Darkness & Pedestrian Gender=Male & Pavement=Dry	0.23	26.83	11.93	1.06
31	Road Type=Rural National & Lighting=Darkness & Pavement=Dry	0.30	23.73	10.55	1.15

ID Rule	Rules with roadway characteristics (Road type) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
32	Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Crash Location Detail=Road section	0.29	25.96	11.54	1.09
33	Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Speed Limit≥60	0.29	25.71	11.43	1.08
34	Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Speed Limit≥60 & Driver Alcohol/Drug use=No	0.29	27.27	12.13	1.06
35	Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Crash Location=Not at intersection	0.29	25.47	11.33	1.07
36	Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.29	27.00	12.00	1.06
37	Road Type=Rural National & Lighting=Darkness & Pavement=Dry & Driver Alcohol/Drug use=No	0.30	25.00	11.12	1.05
38	Road Type=Rural National & Lighting=Darkness & Speed Limit≥60	0.30	22.05	9.80	1.07
39	Road Type=Rural National & Lighting=Darkness & Speed Limit≥60 & Driver Alcohol/Drug use=No	0.30	23.73	10.55	1.08
40	Road Type=Rural National & Lighting=Darkness & Speed Limit≥60 & Crash Location Detail=Road section	0.28	23.64	10.51	1.07
41	Road Type=Rural National & Lighting=Darkness & Speed Limit≥60 & Crash Location=Not at intersection	0.28	23.42	10.41	1.06
42	Road Type=Rural National & Lighting=Darkness & Driver Alcohol/Drug use=No	0.31	21.97	9.77	1.07
43	Road Type=Rural National & Lighting=Darkness & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.29	23.28	10.35	1.06
44	Road Type=Rural National & Lighting=Darkness & Crash Location Detail=Road section	0.29	21.95	9.76	1.07
45	Road Type=Rural National & Lighting=Darkness & Crash Location=Not at intersection	0.29	21.60	9.60	1.05
46	Road Type=Rural National & Driver Gender=Male	0.38	19.57	8.70	1.53
47	Road Type=Rural National & Driver Gender=Male & Pedestrian Gender=Male	0.28	24.53	10.91	1.25
48	Road Type=Rural National & Driver Gender=Male & Pedestrian Gender=Male & Day of Week=Weekday	0.22	28.77	12.79	1.17
49	Road Type=Rural National & Driver Gender=Male & Pedestrian Gender=Male & Speed Limit≥60	0.27	26.88	11.95	1.10
50	Road Type=Rural National & Driver Gender=Male & Pedestrian Gender=Male & Pavement=Dry	0.27	26.32	11.70	1.07
51	Road Type=Rural National & Driver Gender=Male & Pedestrian Gender=Male & N. vehicle involved=1	0.27	26.04	11.58	1.06
52	Road Type=Rural National & Driver Gender=Male & Day of Week=Weekday	0.32	23.26	10.34	1.19
53	Road Type=Rural National & Driver Gender=Male & Day of Week=Weekday & Speed Limit≥60	0.32	25.64	11.40	1.10
54	Road Type=Rural National & Driver Gender=Male & Day of Week=Weekday & Speed Limit≥60 & Driver Gender=Male	0.22	32.31	14.36	1.26
55	Road Type=Rural National & Driver Gender=Male & Day of Week=Weekday & Speed Limit≥60 & Pedestrian Gender=Male	0.30	27.18	12.09	1.06
56	Road Type=Rural National & Driver Gender=Male & Day of Week=Weekday & Pavement=Dry	0.30	24.78	11.02	1.07
57	Road Type=Rural National & Driver Gender=Male & Day of Week=Weekday & Pavement=Dry & Pedestrian Gender=Male	0.21	30.77	13.68	1.24
58	Road Type=Rural National & Driver Gender=Male & Speed Limit≥60	0.37	21.47	9.55	1.10
59	Road Type=Rural National & Driver Gender=Male & Speed Limit≥60 & Pavement=Dry	0.34	22.70	10.09	1.06
60	Road Type=Rural National & Driver Gender=Male & Speed Limit≥60 & Pavement=Dry & Pedestrian Gender=Male	0.25	28.57	12.70	1.26
61	Road Type=Rural National & Driver Gender=Male & Pavement=Dry	0.35	20.63	9.17	1.05
62	Road Type=Rural National & Driver Gender=Male & Pavement=Dry & Crash Location Detail=Road section	0.31	21.97	9.77	1.07
63	Road Type=Rural National & Driver Gender=Male & Pavement=Dry & Crash Location Detail=Road section & Pedestrian Gender=Male	0.22	25.30	11.25	1.15
64	Road Type=Rural National & Pedestrian Gender=Male	0.34	15.46	6.87	1.21
65	Road Type=Rural National & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.34	17.39	7.73	1.13
66	Road Type=Rural National & Pedestrian Gender=Male & Driver Alcohol/Drug use=No & Speed Limit≥60	0.32	18.99	8.44	1.09
67	Road Type=Rural National & Pedestrian Gender=Male & Driver Alcohol/Drug use=No & Pavement=Dry	0.33	18.79	8.35	1.08
68	Road Type=Rural National & Pedestrian Gender=Male & Day of Week=Weekday	0.28	17.11	7.61	1.11

ID Rule	Rules with roadway characteristics (Road type) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
69	Road Type=Rural National & Pedestrian Gender=Male & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.28	19.26	8.56	1.13
70	Road Type=Rural National & Pedestrian Gender=Male & Day of Week=Weekday & Speed Limit≥60	0.27	18.80	8.36	1.10
71	Road Type=Rural National & Pedestrian Gender=Male & Day of Week=Weekday & Pavement=Dry	0.27	18.52	8.23	1.08
72	Road Type=Rural National & Pedestrian Gender=Male & Pavement=Dry	0.33	16.76	7.45	1.08
73	Road Type=Rural National & Pedestrian Gender=Male & Pavement=Dry & Speed Limit≥60	0.31	18.13	8.06	1.08
74	Road Type=Rural National & Pedestrian Gender=Male & Speed Limit≥60	0.32	16.67	7.41	1.08
75	Road Type=Rural National & Speed Limit≥60	0.46	14.29	6.35	1.11
76	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday	0.37	16.13	7.17	1.13
77	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.37	17.68	7.86	1.10
78	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.27	21.37	9.50	1.21
79	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Pavement=Dry	0.35	17.46	7.76	1.08
80	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Pavement=Dry & Pedestrian Gender=Male	0.25	20.34	9.04	1.16
81	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Pavement=Dry & Driver Alcohol/Drug use=No	0.35	18.97	8.43	1.09
82	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Crash Location Detail=Road section	0.32	17.14	7.62	1.06
83	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Crash Location Detail=Road section & Driver Gender=Male	0.27	27.17	12.08	1.59
84	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Crash Location Detail=Road section & Pedestrian Gender=Male	0.21	18.69	8.31	1.09
85	Road Type=Rural National & Speed Limit≥60 & Day of Week=Weekday & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.32	18.87	8.39	1.10
86	Road Type=Rural National & Speed Limit≥60 & Driver Alcohol/Drug use=No	0.45	15.56	6.92	1.09
87	Road Type=Rural National & Speed Limit≥60 & Driver Alcohol/Drug use=No & Pavement=Dry	0.41	16.67	7.41	1.07
88	Road Type=Rural National & Speed Limit≥60 & Driver Alcohol/Drug use=No & Pavement=Dry & Pedestrian Gender=Male	0.31	20.57	9.14	1.23
89	Road Type=Rural National & Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.39	16.52	7.34	1.06
90	Road Type=Rural National & Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location Detail=Road section & Pedestrian Gender=Male	0.27	18.94	8.42	1.15
91	Road Type=Rural National & Speed Limit≥60 & Pavement=Dry	0.42	15.33	6.81	1.07
92	Road Type=Rural National & Speed Limit≥60 & Pavement=Dry & Crash Location Detail=Road section	0.38	16.59	7.38	1.08
93	Road Type=Rural National & Speed Limit≥60 & Pavement=Dry & Crash Location Detail=Road section & Driver Gender=Male	0.30	24.14	10.73	1.45
94	Road Type=Rural National & Speed Limit≥60 & Pavement=Dry & Crash Location Detail=Road section & Pedestrian Gender=Male	0.27	18.52	8.23	1.12
95	Road Type=Rural National & Speed Limit≥60 & Pavement=Dry & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.37	18.23	8.11	1.10
96	Road Type=Rural National & Speed Limit≥60 & Crash Location Detail=Road section	0.40	15.14	6.73	1.06
97	Road Type=Rural National & Day of Week=Weekday	0.38	14.29	6.35	1.11
98	Road Type=Rural National & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.38	15.58	6.93	1.09
99	Road Type=Rural National & Day of Week=Weekday & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.33	16.94	7.53	1.09
100	Road Type=Rural National & Day of Week=Weekday & Driver Alcohol/Drug use=No & Crash Location Detail=Road section & Pedestrian Gender=Male	0.22	18.92	8.41	1.12
101	Road Type=Rural National & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pavement=Dry	0.36	16.59	7.37	1.06
102	Road Type=Rural National & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pavement=Dry & Pedestrian Gender=Male	0.27	20.66	9.19	1.25
103	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section	0.33	15.42	6.86	1.08
104	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry	0.32	17.05	7.58	1.11
105	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry & Speed Limit≥60	0.31	19.08	8.48	1.12

ID Rule	Rules with roadway characteristics (Road type) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
106	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry & Lighting=Darkness	0.23	33.85	15.05	1.99
107	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry & Driver Gender=Male	0.25	26.67	11.86	1.56
108	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry & Pedestrian Gender=Male	0.22	18.58	8.26	1.09
109	Road Type=Rural National & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry & Driver Alcohol/Drug use=No	0.32	18.75	8.34	1.10
110	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry	0.36	15.32	6.81	1.07
111	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry & Crash Location=Not at intersection	0.32	16.22	7.21	1.06
112	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry & Crash Location=Not at intersection & Speed Limit≥60	0.31	18.24	8.11	1.12
113	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry & Crash Location=Not at intersection & Lighting=Darkness	0.23	33.33	14.82	2.06
114	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry & Crash Location=Not at intersection & Driver Gender=Male	0.25	26.09	11.60	1.61
115	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry & Crash Location=Not at intersection & Pedestrian Gender=Male	0.22	18.10	8.05	1.12
116	Road Type=Rural National & Day of Week=Weekday & Pavement=Dry & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.32	17.75	7.89	1.09
117	Road Type=Rural National & Crash Location Detail=Road section	0.42	13.94	6.20	1.09
118	Road Type=Rural National & Crash Location Detail=Road section & Pavement=Dry	0.40	15.14	6.73	1.09
119	Road Type=Rural National & Crash Location Detail=Road section & Pavement=Dry & Driver Alcohol/Drug use=No	0.39	16.52	7.34	1.09
120	Road Type=Rural National & Crash Location Detail=Road section & Pavement=Dry & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.29	19.42	8.64	1.18
121	Road Type=Rural National & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.41	15.12	6.72	1.08
122	Road Type=Rural National & Driver Alcohol/Drug use=No	0.47	13.84	6.15	1.08
123	Road Type=Rural National & Driver Alcohol/Drug use=No & Pavement=Dry	0.43	14.64	6.51	1.06
124	Road Type=Rural National & Driver Alcohol/Drug use=No & Pavement=Dry & Crash Location=Not at intersection	0.39	15.61	6.94	1.07
125	Road Type=Rural National & Driver Alcohol/Drug use=No & Pavement=Dry & Crash Location=Not at intersection & Pedestrian Gender=Male	0.29	18.88	8.39	1.21
126	Road Type=Rural National & Pavement=Dry	0.45	13.59	6.04	1.06
127	Road Type=Rural National & Pavement=Dry & Crash Location=Not at intersection	0.40	14.39	6.40	1.06

Table 159 – Association rules with roadway characteristics (Speed limit) as first antecedent and fatal crashes as consequent, Sweden.

ID Rule	Rules with roadway characteristics (Speed limit) as first antecedent and fatal pedestrian crash as consequent Antecedents	S %	C %	Lift	LIC
128	Speed Limit≥60	0.82	10.49	4.66	n.a.
129	Speed Limit≥60 & Vehicle Type=Truck	0.29	22.88	10.17	2.18
130	Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male	0.28	28.57	12.70	1.25
131	Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male & Area=Rural	0.20	32.20	14.32	1.13
132	Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male & Pedestrian Gender=Male	0.20	30.65	13.63	1.07
133	Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male & Crash Location Detail=Road section	0.23	30.56	13.59	1.07
134	Speed Limit≥60 & Vehicle Type=Truck & Driver Gender=Male & Crash Location=Not at intersection	0.23	30.14	13.40	1.05
135	Speed Limit≥60 & Vehicle Type=Truck & Area=Rural	0.21	27.03	12.02	1.18
136	Speed Limit≥60 & Vehicle Type=Truck & Pedestrian Gender=Male	0.21	25.32	11.26	1.11
137	Speed Limit≥60 & Vehicle Type=Truck & Pedestrian Gender=Male & Pavement=Dry	0.21	26.67	11.86	1.05
138	Speed Limit≥60 & Vehicle Type=Truck & Crash Location Detail=Road section	0.24	25.00	11.12	1.09
139	Speed Limit≥60 & Vehicle Type=Truck & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.24	26.44	11.75	1.06
140	Speed Limit≥60 & Vehicle Type=Truck & Crash Location Detail=Road section & Pavement=Dry	0.24	26.44	11.75	1.06
141	Speed Limit≥60 & Vehicle Type=Truck & Crash Location Detail=Road section & Pavement=Dry & Driver Gender=Male	0.23	31.43	13.97	1.19
142	Speed Limit≥60 & Vehicle Type=Truck & Crash Location=Not at intersection	0.24	24.21	10.76	1.06
143	Speed Limit≥60 & Vehicle Type=Truck & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.24	25.56	11.36	1.06
144	Speed Limit≥60 & Vehicle Type=Truck & Crash Location=Not at intersection & Pavement=Dry	0.24	25.56	11.36	1.06
145	Speed Limit≥60 & Vehicle Type=Truck & Crash Location=Not at intersection & Pavement=Dry & Driver Gender=Male	0.23	30.99	13.78	1.21
146	Speed Limit≥60 & Vehicle Type=Truck & Pavement=Dry	0.29	24.11	10.72	1.05
147	Speed Limit≥60 & Driver Gender=Male	0.63	16.08	7.15	1.53
148	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness	0.32	21.43	9.53	1.33
149	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Area=Rural	0.29	28.13	12.51	1.31
150	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Area=Rural & Pedestrian Gender=Male	0.21	32.26	14.34	1.15
151	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Area=Rural & N. Pedestrian involved=1	0.28	29.89	13.29	1.06
152	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Day of Week=Weekday	0.28	27.96	12.43	1.30
153	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Pedestrian Gender=Male	0.22	24.42	10.86	1.14
154	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Pavement=Dry	0.30	23.53	10.46	1.10
155	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Pavement=Dry & Pedestrian Gender=Male	0.21	25.00	11.12	1.06
156	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section	0.28	23.01	10.23	1.07
157	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Crash Location=Not at intersection	0.28	23.01	10.23	1.07
158	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section & Pedestrian Gender=Male	0.20	25.33	11.26	1.10
159	Speed Limit≥60 & Driver Gender=Male & Lighting=Darkness & Crash Location=Not at intersection & Pedestrian Gender=Male	0.20	25.33	11.26	1.10
160	Speed Limit≥60 & Driver Gender=Male & Area=Rural	0.49	20.09	8.93	1.25
161	Speed Limit≥60 & Driver Gender=Male & Area=Rural & Pedestrian Gender=Male	0.35	24.26	10.79	1.21
162	Speed Limit≥60 & Driver Gender=Male & Area=Rural & Day of Week=Weekday	0.41	24.22	10.77	1.21
163	Speed Limit≥60 & Driver Gender=Male & Area=Rural & Day of Week=Weekday & Pedestrian Gender=Male	0.30	29.47	13.10	1.22

ID Rule	Rules with roadway characteristics (Speed limit) as first antecedent and fatal pedestrian crash as consequent Antecedents	S %	C %	Lift	LIC
164	Speed Limit≥60 & Driver Gender=Male & Area=Rural & Road Type=Rural National	0.37	21.47	9.55	1.07
165	Speed Limit≥60 & Driver Gender=Male & Area=Rural & Road Type=Rural National & Pedestrian Gender=Male	0.27	26.88	11.95	1.25
166	Speed Limit≥60 & Driver Gender=Male & Crash Location Detail=Road section	0.54	17.89	7.96	1.11
167	Speed Limit≥60 & Driver Gender=Male & Crash Location Detail=Road section & Pavement=Dry	0.52	19.37	8.61	1.08
168	Speed Limit≥60 & Driver Gender=Male & Crash Location Detail=Road section & Day of Week=Weekday	0.42	19.23	8.55	1.07
169	Speed Limit≥60 & Driver Gender=Male & Crash Location Detail=Road section & Day of Week=Weekday & Pedestrian Gender=Male	0.29	20.61	9.16	1.07
170	Speed Limit≥60 & Driver Gender=Male & Pedestrian Gender=Male	0.40	17.59	7.82	1.09
171	Speed Limit≥60 & Driver Gender=Male & Pedestrian Gender=Male & Day of Week=Weekday	0.34	19.39	8.62	1.10
172	Speed Limit≥60 & Driver Gender=Male & Pedestrian Gender=Male & Pavement=Dry	0.39	18.69	8.31	1.06
173	Speed Limit≥60 & Driver Gender=Male & Pedestrian Gender=Male & N. Pedestrian involved=1	0.40	18.63	8.28	1.06
174	Speed Limit≥60 & Driver Gender=Male & Day of Week=Weekday	0.51	17.39	7.73	1.08
175	Speed Limit≥60 & Driver Gender=Male & Day of Week=Weekday & Crash Location=Not at intersection	0.42	18.26	8.12	1.05
176	Speed Limit≥60 & Driver Gender=Male & Day of Week=Weekday & Crash Location=Not at intersection & Pedestrian Gender=Male	0.29	20.00	8.89	1.10
177	Speed Limit≥60 & Driver Gender=Male & Crash Location=Not at intersection	0.54	17.17	7.63	1.07
178	Speed Limit≥60 & Driver Gender=Male & Crash Location=Not at intersection & Pavement=Dry	0.52	18.49	8.22	1.08
179	Speed Limit≥60 & Driver Gender=Male & Crash Location=Not at intersection & Pavement=Dry & Pedestrian Gender=Male	0.35	19.76	8.79	1.07
180	Speed Limit≥60 & Lighting=Darkness	0.43	14.80	6.58	1.41
181	Speed Limit≥60 & Lighting=Darkness & Area=Rural	0.38	20.45	9.09	1.38
182	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Day of Week=Weekday	0.32	25.64	11.40	1.49
183	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Day of Week=Weekday & Driver Gender=Male	0.25	41.38	18.40	1.61
184	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Day of Week=Weekday & Pedestrian Gender=Male	0.23	30.14	13.40	1.18
185	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.32	27.52	12.24	1.07
186	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pedestrian Gender=Male	0.29	24.55	10.91	1.20
187	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pavement=Dry	0.36	23.45	10.43	1.15
188	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pavement=Dry & Driver Gender=Male	0.27	30.49	13.56	1.30
189	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pavement=Dry & Pedestrian Gender=Male	0.28	27.08	12.04	1.16
190	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Pavement=Dry & Driver Alcohol/Drug use=No	0.36	25.00	11.12	1.07
191	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Driver Alcohol/Drug use=No	0.38	22.09	9.82	1.08
192	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.29	27.27	12.13	1.23
193	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location Detail=Road section	0.35	21.85	9.72	1.07
194	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location Detail=Road section & Driver Gender=Male	0.25	28.57	12.70	1.31
195	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location Detail=Road section & Pedestrian Gender=Male	0.27	25.51	11.34	1.17
196	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.35	23.57	10.48	1.08
197	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location=Not at intersection	0.35	21.57	9.59	1.05
198	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location=Not at intersection & Driver Gender=Male	0.25	28.57	12.70	1.32
199	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location=Not at intersection & Pedestrian Gender=Male	0.27	25.00	11.12	1.16
200	Speed Limit≥60 & Lighting=Darkness & Area=Rural & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.35	23.24	10.33	1.08

ID Rule	Rules with roadway characteristics (Speed limit) as first antecedent and fatal pedestrian crash as consequent Antecedents	S %	C %	Lift	LIC
201	Speed Limit≥60 & Lighting=Darkness & Pedestrian Gender=Male	0.31	17.68	7.86	1.19
202	Speed Limit≥60 & Lighting=Darkness & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.31	20.42	9.08	1.15
203	Speed Limit≥60 & Lighting=Darkness & Day of Week=Weekday	0.36	17.26	7.67	1.17
204	Speed Limit≥60 & Lighting=Darkness & Day of Week=Weekday & Pedestrian Gender=Male	0.25	20.34	9.04	1.18
205	Speed Limit≥60 & Lighting=Darkness & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.35	18.86	8.38	1.09
206	Speed Limit≥60 & Lighting=Darkness & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.25	23.08	10.26	1.22
207	Speed Limit≥60 & Lighting=Darkness & Crash Location Detail=Road section	0.38	16.67	7.41	1.13
208	Speed Limit≥60 & Lighting=Darkness & Crash Location Detail=Road section & Pedestrian Gender=Male	0.29	20.00	8.89	1.20
209	Speed Limit≥60 & Lighting=Darkness & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.38	18.65	8.29	1.12
210	Speed Limit≥60 & Lighting=Darkness & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.29	22.69	10.09	1.22
211	Speed Limit≥60 & Lighting=Darkness & Pavement=Dry	0.40	16.59	7.38	1.12
212	Speed Limit≥60 & Lighting=Darkness & Pavement=Dry & Pedestrian Gender=Male	0.29	19.15	8.51	1.15
213	Speed Limit≥60 & Lighting=Darkness & Pavement=Dry & Driver Alcohol/Drug use=No	0.39	18.32	8.14	1.10
214	Speed Limit≥60 & Lighting=Darkness & Pavement=Dry & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.29	21.95	9.76	1.20
215	Speed Limit≥60 & Lighting=Darkness & Driver Alcohol/Drug use=No	0.42	16.39	7.29	1.11
216	Speed Limit≥60 & Lighting=Darkness & Crash Location=Not at intersection	0.38	16.36	7.28	1.11
217	Speed Limit≥60 & Lighting=Darkness & Crash Location=Not at intersection & Pedestrian Gender=Male	0.29	19.57	8.70	1.20
218	Speed Limit≥60 & Lighting=Darkness & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.38	18.27	8.13	1.12
219	Speed Limit≥60 & Lighting=Darkness & Crash Location=Not at intersection & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.29	22.13	9.84	1.21
220	Speed Limit≥60 & Area=Rural	0.64	13.99	6.22	1.33
221	Speed Limit≥60 & Area=Rural & Pedestrian Gender=Male	0.45	16.22	7.21	1.16
222	Speed Limit≥60 & Area=Rural & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.45	18.10	8.05	1.12
223	Speed Limit≥60 & Area=Rural & Pedestrian Gender=Male & Pavement=Dry	0.43	17.83	7.93	1.10
224	Speed Limit≥60 & Area=Rural & Pedestrian Gender=Male & Crash Location Detail=Road section	0.39	17.13	7.62	1.06
225	Speed Limit≥60 & Area=Rural & Day of Week=Weekday	0.51	15.95	7.09	1.14
226	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Pedestrian Gender=Male	0.37	18.82	8.37	1.18
227	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.51	17.39	7.73	1.09
228	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.37	21.08	9.37	1.21
229	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Crash Location Detail=Road section	0.43	17.15	7.63	1.08
230	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Crash Location Detail=Road section & Driver Gender=Male	0.35	25.78	11.46	1.50
231	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Crash Location Detail=Road section & Pedestrian Gender=Male	0.32	19.74	8.78	1.15
232	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.43	18.81	8.36	1.10
233	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Crash Location Detail=Road section & Pavement=Dry	0.42	18.87	8.39	1.10
234	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Pavement=Dry	0.48	16.92	7.52	1.06
235	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Pavement=Dry & Lighting=Darkness	0.30	29.17	12.97	1.72
236	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Pavement=Dry & Driver Gender=Male	0.38	25.00	11.12	1.48
237	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Pavement=Dry & Pedestrian Gender=Male	0.36	20.36	9.05	1.20

ID Rule	Rules with roadway characteristics (Speed limit) as first antecedent and fatal pedestrian crash as consequent Antecedents	S %	C %	Lift	LIC
238	Speed Limit≥60 & Area=Rural & Day of Week=Weekday & Pavement=Dry & Driver Alcohol/Drug use=No	0.48	18.37	8.17	1.09
239	Speed Limit≥60 & Area=Rural & Driver Alcohol/Drug use=No	0.63	15.13	6.73	1.08
240	Speed Limit≥60 & Area=Rural & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.55	16.30	7.25	1.08
241	Speed Limit≥60 & Area=Rural & Driver Alcohol/Drug use=No & Crash Location Detail=Road section & Pedestrian Gender=Male	0.39	19.17	8.52	1.18
242	Speed Limit≥60 & Area=Rural & Driver Alcohol/Drug use=No & Pavement=Dry	0.58	16.08	7.15	1.06
243	Speed Limit≥60 & Area=Rural & Driver Alcohol/Drug use=No & Pavement=Dry & Pedestrian Gender=Male	0.43	19.81	8.81	1.23
244	Speed Limit≥60 & Crash Location Detail=Road section	0.70	12.27	5.45	1.17
245	Speed Limit≥60 & Crash Location Detail=Road section & Area=Rural	0.56	15.01	6.68	1.22
246	Speed Limit≥60 & Crash Location Detail=Road section & Area=Rural & Pavement=Dry	0.54	16.45	7.31	1.10
247	Speed Limit≥60 & Crash Location Detail=Road section & Area=Rural & Pavement=Dry & Driver Gender=Male	0.40	22.75	10.12	1.38
248	Speed Limit≥60 & Crash Location Detail=Road section & Area=Rural & Pavement=Dry & Pedestrian Gender=Male	0.39	19.07	8.48	1.16
249	Speed Limit≥60 & Crash Location Detail=Road section & Area=Rural & Pavement=Dry & Driver Alcohol/Drug use=No	0.53	17.92	7.97	1.09
250	Speed Limit≥60 & Crash Location Detail=Road section & Pedestrian Gender=Male	0.47	13.62	6.06	1.11
251	Speed Limit≥60 & Crash Location Detail=Road section & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.47	15.38	6.84	1.13
252	Speed Limit≥60 & Crash Location Detail=Road section & Pedestrian Gender=Male & Pavement=Dry	0.46	15.03	6.68	1.10
253	Speed Limit≥60 & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.69	13.49	6.00	1.10
254	Speed Limit≥60 & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pavement=Dry	0.66	14.66	6.52	1.09
255	Speed Limit≥60 & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pavement=Dry & Pedestrian Gender=Male	0.46	17.00	7.56	1.16
256	Speed Limit≥60 & Pedestrian Gender=Male	0.52	11.53	5.13	1.10
257	Speed Limit≥60 & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.52	13.10	5.83	1.14
258	Speed Limit≥60 & Pedestrian Gender=Male & Crash Location=Not at intersection	0.47	12.75	5.67	1.11
259	Speed Limit≥60 & Pedestrian Gender=Male & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.47	14.33	6.37	1.12
260	Speed Limit≥60 & Pedestrian Gender=Male & Crash Location=Not at intersection & Pavement=Dry	0.46	14.14	6.29	1.11
261	Speed Limit≥60 & Pedestrian Gender=Male & Crash Location=Not at intersection & Crash Location Detail=Road section	0.47	13.62	6.06	1.07
262	Speed Limit≥60 & Pedestrian Gender=Male & Crash Location=Not at intersection & Day of Week=Weekday	0.38	13.58	6.04	1.07
263	Speed Limit≥60 & Pedestrian Gender=Male & Pavement=Dry	0.50	12.53	5.57	1.09
264	Speed Limit≥60 & Pedestrian Gender=Male & Pavement=Dry & Driver Alcohol/Drug use=No	0.50	14.24	6.33	1.14
265	Speed Limit≥60 & Pedestrian Gender=Male & Pavement=Dry & Day of Week=Weekday	0.41	13.40	5.96	1.07
266	Speed Limit≥60 & Pedestrian Gender=Male & Day of Week=Weekday	0.43	12.46	5.54	1.08
267	Speed Limit≥60 & Pedestrian Gender=Male & Day of Week=Weekday & Crash Location Detail=Road section	0.38	14.57	6.48	1.17
268	Speed Limit≥60 & Pedestrian Gender=Male & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.43	14.04	6.24	1.13
269	Speed Limit≥60 & Driver Alcohol/Drug use=No	0.80	11.50	5.11	1.10
270	Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location=Not at intersection	0.69	12.29	5.46	1.07
271	Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location=Not at intersection & Crash Location Detail=Road section	0.69	13.49	6.00	1.10
272	Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location=Not at intersection & Crash Location Detail=Road section & Pedestrian Gender=Male	0.47	15.38	6.84	1.14
273	Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location=Not at intersection & Pavement=Dry	0.66	13.33	5.93	1.09
274	Speed Limit≥60 & Driver Alcohol/Drug use=No & Crash Location=Not at intersection & Pavement=Dry & Pedestrian Gender=Male	0.46	15.93	7.08	1.19

ID Rule	Rules with roadway characteristics (Speed limit) as first antecedent and fatal pedestrian crash as consequent Antecedents	S %	C %	Lift	LIC
275	Speed Limit≥60 & Driver Alcohol/Drug use=No & Pavement=Dry	0.73	12.08	5.37	1.05
276	Speed Limit≥60 & Crash Location=Not at intersection	0.70	11.21	4.98	1.07
277	Speed Limit≥60 & Crash Location=Not at intersection & Area=Rural	0.56	14.21	6.32	1.27
278	Speed Limit≥60 & Crash Location=Not at intersection & Area=Rural & Pavement=Dry	0.54	15.55	6.91	1.09
279	Speed Limit≥60 & Crash Location=Not at intersection & Area=Rural & Pavement=Dry & Lighting=Darkness	0.35	25.78	11.46	1.66
280	Speed Limit≥60 & Crash Location=Not at intersection & Area=Rural & Pavement=Dry & Driver Gender=Male	0.40	21.84	9.71	1.40
281	Speed Limit≥60 & Crash Location=Not at intersection & Area=Rural & Pavement=Dry & Pedestrian Gender=Male	0.39	18.32	8.14	1.18
282	Speed Limit≥60 & Crash Location=Not at intersection & Area=Rural & Pavement=Dry & Driver Alcohol/Drug use=No	0.53	16.89	7.51	1.09

Table 160 – Association rules with roadway characteristics (Area) as first antecedent and fatal crashes as consequent, Sweden.

ID Rule	Rules with roadway characteristics (Area) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
283	Area=Rural	0.79	7.23	3.22	n.a.
284	Area=Rural & Vehicle Type=Truck	0.23	15.71	6.99	2.17
285	Area=Rural & Vehicle Type=Truck & Driver Gender=Male	0.21	22.22	9.88	1.41
286	Area=Rural & Vehicle Type=Truck & Driver Gender=Male & Pavement=Dry	0.21	23.53	10.46	1.06
287	Area=Rural & Vehicle Type=Truck & Crash Location Detail=Road section	0.20	18.27	8.12	1.16
288	Area=Rural & Vehicle Type=Truck & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.20	19.59	8.71	1.07
289	Area=Rural & Vehicle Type=Truck & Crash Location=Not at intersection	0.20	17.27	7.68	1.10
290	Area=Rural & Vehicle Type=Truck & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.20	18.45	8.20	1.07
291	Area=Rural & Vehicle Type=Truck & Driver Alcohol/Drug use=No	0.23	16.79	7.47	1.07
292	Area=Rural & Vehicle Type=Truck & Day of Week=Weekday	0.21	16.53	7.35	1.05
293	Area=Rural & Driver Gender=Male	0.58	12.33	5.48	1.70
294	Area=Rural & Driver Gender=Male & Lighting=Darkness	0.33	21.38	9.51	1.73
295	Area=Rural & Driver Gender=Male & Lighting=Darkness & Pedestrian Gender=Male	0.25	27.59	12.27	1.29
296	Area=Rural & Driver Gender=Male & Lighting=Darkness & Pavement=Dry	0.31	23.02	10.23	1.08
297	Area=Rural & Driver Gender=Male & Lighting=Darkness & Pavement=Dry & Pedestrian Gender=Male	0.24	28.75	12.78	1.25
298	Area=Rural & Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section	0.30	22.58	10.04	1.06
299	Area=Rural & Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section & Pedestrian Gender=Male	0.23	27.16	12.08	1.20
300	Area=Rural & Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section & N. Pedestrian involved=1	0.29	23.89	10.62	1.06
301	Area=Rural & Driver Gender=Male & Lighting=Darkness & N. Pedestrian involved=1	0.32	22.56	10.03	1.06
302	Area=Rural & Driver Gender=Male & Lighting=Darkness & N. Pedestrian involved=1 & Pedestrian Gender=Male	0.25	29.27	13.01	1.30
303	Area=Rural & Driver Gender=Male & Pedestrian Gender=Male	0.42	17.17	7.63	1.39
304	Area=Rural & Driver Gender=Male & Pedestrian Gender=Male & Day of Week=Weekday	0.35	20.37	9.06	1.19
305	Area=Rural & Driver Gender=Male & Pedestrian Gender=Male & Pavement=Dry	0.41	18.22	8.10	1.06

ID Rule	Rules with roadway characteristics (Area) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
306	Area=Rural & Driver Gender=Male & Pedestrian Gender=Male & N. Pedestrian involved=1	0.42	18.10	8.05	1.05
307	Area=Rural & Driver Gender=Male & Day of Week=Weekday	0.49	14.51	6.45	1.18
308	Area=Rural & Driver Gender=Male & Day of Week=Weekday & Crash Location Detail=Road section	0.42	17.02	7.57	1.17
309	Area=Rural & Driver Gender=Male & Day of Week=Weekday & Crash Location Detail=Road section & Pedestrian Gender=Male	0.30	20.74	9.22	1.22
310	Area=Rural & Driver Gender=Male & Crash Location Detail=Road section	0.52	14.24	6.33	1.16
311	Area=Rural & Driver Gender=Male & Crash Location Detail=Road section & Pavement=Dry	0.50	15.21	6.76	1.07
312	Area=Rural & Driver Gender=Male & Crash Location Detail=Road section & Pavement=Dry & Pedestrian Gender=Male	0.37	19.02	8.46	1.25
313	Area=Rural & Lighting=Darkness	0.43	12.31	5.47	1.70
314	Area=Rural & Lighting=Darkness & Pedestrian Gender=Male	0.34	15.61	6.94	1.27
315	Area=Rural & Lighting=Darkness & Pedestrian Gender=Male & Day of Week=Weekday	0.28	19.40	8.63	1.24
316	Area=Rural & Lighting=Darkness & Pedestrian Gender=Male & Crash Location Detail=Road section	0.32	17.86	7.94	1.14
317	Area=Rural & Lighting=Darkness & Pedestrian Gender=Male & Pavement=Dry	0.33	17.42	7.74	1.12
318	Area=Rural & Lighting=Darkness & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.34	17.30	7.69	1.11
319	Area=Rural & Lighting=Darkness & Day of Week=Weekday	0.36	15.18	6.75	1.23
320	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Driver Gender=Male	0.29	30.34	13.49	2.00
321	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Driver Gender=Male & Pedestrian Gender=Male	0.21	39.22	17.44	1.29
322	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Driver Gender=Male & Crash Location Detail=Road section	0.25	33.33	14.82	1.10
323	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Driver Gender=Male & Pavement=Dry	0.27	32.05	14.25	1.06
324	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location Detail=Road section	0.33	19.14	8.51	1.26
325	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location Detail=Road section & Pedestrian Gender=Male	0.25	22.22	9.88	1.16
326	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.33	20.53	9.13	1.07
327	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Pavement=Dry	0.34	16.84	7.49	1.11
328	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Pavement=Dry & Pedestrian Gender=Male	0.27	21.55	9.58	1.28
329	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Pavement=Dry & Driver Alcohol/Drug use=No	0.34	18.39	8.18	1.09
330	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location=Not at intersection	0.33	16.67	7.41	1.10
331	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location=Not at intersection & Speed Limit≥60	0.29	27.27	12.13	1.64
332	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location=Not at intersection & Pedestrian Gender=Male	0.25	20.69	9.20	1.24
333	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.33	17.92	7.97	1.08
334	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.36	16.59	7.37	1.09
335	Area=Rural & Lighting=Darkness & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.28	21.67	9.63	1.31
336	Area=Rural & Lighting=Darkness & Crash Location Detail=Road section	0.40	15.02	6.68	1.22
337	Area=Rural & Lighting=Darkness & Crash Location Detail=Road section & Pavement=Dry	0.40	17.67	7.86	1.18
338	Area=Rural & Lighting=Darkness & Crash Location Detail=Road section & Pavement=Dry & Pedestrian Gender=Male	0.32	20.13	8.95	1.14
339	Area=Rural & Lighting=Darkness & Crash Location Detail=Road section & Pavement=Dry & Driver Alcohol/Drug use=No	0.40	19.19	8.53	1.09
340	Area=Rural & Lighting=Darkness & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.40	16.24	7.22	1.08
341	Area=Rural & Lighting=Darkness & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.32	19.48	8.66	1.20
342	Area=Rural & Lighting=Darkness & Pavement=Dry	0.41	13.83	6.15	1.12

ID Rule	Rules with roadway characteristics (Area) as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
343	Area=Rural & Lighting=Darkness & Pavement=Dry & Crash Location=Not at intersection	0.40	15.51	6.90	1.12
344	Area=Rural & Lighting=Darkness & Pavement=Dry & Crash Location=Not at intersection & Driver Gender=Male	0.30	25.00	11.12	1.61
345	Area=Rural & Lighting=Darkness & Pavement=Dry & Crash Location=Not at intersection & Pedestrian Gender=Male	0.32	18.40	8.18	1.19
346	Area=Rural & Lighting=Darkness & Pavement=Dry & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.40	16.81	7.48	1.08
347	Area=Rural & Lighting=Darkness & Pavement=Dry & Driver Alcohol/Drug use=No	0.41	15.06	6.70	1.09
348	Area=Rural & Lighting=Darkness & Pavement=Dry & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.33	19.25	8.56	1.28
349	Area=Rural & Lighting=Darkness & Driver Alcohol/Drug use=No	0.43	13.40	5.96	1.09
350	Area=Rural & Lighting=Darkness & Driver Alcohol/Drug use=No & Crash Location=Not at intersection	0.40	14.29	6.35	1.07
351	Area=Rural & Lighting=Darkness & Driver Alcohol/Drug use=No & Crash Location=Not at intersection & Pedestrian Gender=Male	0.32	17.65	7.85	1.24
352	Area=Rural & Lighting=Darkness & Crash Location=Not at intersection	0.40	13.24	5.89	1.08
353	Area=Rural & Pedestrian Gender=Male	0.55	9.68	4.31	1.34
354	Area=Rural & Pedestrian Gender=Male & Day of Week=Weekday	0.46	11.47	5.10	1.18
355	Area=Rural & Pedestrian Gender=Male & Day of Week=Weekday & Crash Location Detail=Road section	0.40	12.97	5.77	1.13
356	Area=Rural & Pedestrian Gender=Male & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.46	12.95	5.76	1.13
357	Area=Rural & Pedestrian Gender=Male & Day of Week=Weekday & Pavement=Dry	0.43	12.06	5.36	1.05
358	Area=Rural & Pedestrian Gender=Male & Crash Location Detail=Road section	0.50	11.30	5.02	1.17
359	Area=Rural & Pedestrian Gender=Male & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.50	12.67	5.63	1.12
360	Area=Rural & Pedestrian Gender=Male & Crash Location Detail=Road section & Pavement=Dry	0.49	12.17	5.41	1.08
361	Area=Rural & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.55	10.79	4.80	1.11
362	Area=Rural & Pedestrian Gender=Male & Driver Alcohol/Drug use=No & Pavement=Dry	0.53	11.60	5.16	1.08
363	Area=Rural & Pedestrian Gender=Male & Pavement=Dry	0.53	10.37	4.61	1.07
364	Area=Rural & Crash Location Detail=Road section	0.71	9.44	4.20	1.30
365	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday	0.56	10.66	4.74	1.13
366	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.56	11.83	5.26	1.11
367	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.40	14.67	6.52	1.24
368	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Pavement=Dry	0.54	11.31	5.03	1.06
369	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Pavement=Dry & Lighting=Darkness	0.33	22.63	10.06	2.00
370	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Pavement=Dry & Driver Gender=Male	0.41	18.22	8.10	1.61
371	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Pavement=Dry & Pedestrian Gender=Male	0.39	13.81	6.14	1.22
372	Area=Rural & Crash Location Detail=Road section & Day of Week=Weekday & Pavement=Dry & Driver Alcohol/Drug use=No	0.54	12.62	5.61	1.12
373	Area=Rural & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.70	10.31	4.59	1.09
374	Area=Rural & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pavement=Dry	0.67	10.98	4.88	1.06
375	Area=Rural & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pavement=Dry & Pedestrian Gender=Male	0.49	13.73	6.11	1.25
376	Area=Rural & Crash Location Detail=Road section & Pavement=Dry	0.68	9.98	4.44	1.06
377	Area=Rural & Day of Week=Weekday	0.64	8.22	3.65	1.14
378	Area=Rural & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.64	9.04	4.02	1.10
379	Area=Rural & Driver Alcohol/Drug use=No	0.77	7.82	3.48	1.08

Table 161 – Association rules with pedestrian characteristics as first antecedent and fatal crashes as consequent, Sweden.

ID Rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
380	Pedestrian Age≥75	0.67	5.92	2.63	n.a.
381	Pedestrian Age≥75 & Driver Gender=Male	0.47	9.32	4.14	1.57
382	Pedestrian Age≥75 & Driver Gender=Male & Crash Location Detail=Road section	0.35	10.44	4.64	1.12
383	Pedestrian Age≥75 & Driver Gender=Male & Crash Location Detail=Road section & Pedestrian Gender=Female	0.23	11.00	4.89	1.05
384	Pedestrian Age≥75 & Driver Gender=Male & Crash Location Detail=Road section & Pedestrian Gender=Female & Area=Urban	0.23	11.96	5.32	1.09
385	Pedestrian Age≥75 & Speed Limit=50	0.27	7.72	3.43	1.30
386	Pedestrian Age≥75 & Crash Location Detail=Road section	0.50	7.44	3.31	1.26
387	Pedestrian Age≥75 & Crash Location Detail=Road section & Pedestrian Gender=Male	0.21	8.58	3.82	1.15
388	Pedestrian Age≥75 & Crash Location Detail=Road section & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.21	9.17	4.08	1.07
389	Pedestrian Age≥75 & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.50	7.91	3.52	1.06
390	Pedestrian Age≥75 & Pedestrian Gender=Male	0.29	7.34	3.26	1.24
391	Pedestrian Age≥75 & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.29	7.87	3.50	1.07
392	Pedestrian Age≥75 & Driver Alcohol/Drug use=No	0.67	6.31	2.80	1.07
393	Pedestrian Age≥75 & Driver Alcohol/Drug use=No & Driver Gender=Male	0.47	9.32	4.14	1.48
394	Pedestrian Age=65-74	0.35	3.91	1.74	n.a.
395	Pedestrian Age=65-74 & Pedestrian Gender=Male	0.22	6.10	2.71	1.56
396	Pedestrian Age=65-74 & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.21	6.78	3.01	1.11
397	Pedestrian Age=65-74 & Crash Location Detail=Road section	0.27	5.22	2.32	1.34
398	Pedestrian Age=65-74 & Vehicle Type=Car	0.24	4.33	1.93	1.11
399	Pedestrian Age=65-74 & Vehicle Type=Car & Crash Location Detail=Road section	0.21	5.65	2.51	1.30
400	Pedestrian Age=65-74 & Vehicle Type=Car & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.21	6.31	2.81	1.12
401	Pedestrian Age=65-74 & Vehicle Type=Car & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & N Pedestrian involved=1	0.21	6.64	2.95	1.05
402	Pedestrian Age=65-74 & Vehicle Type=Car & Crash Location=Not at intersection	0.21	4.98	2.21	1.15
403	Pedestrian Age=65-74 & Vehicle Type=Car & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.21	5.52	2.46	1.11
404	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No	0.34	4.20	1.87	1.08
405	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.25	5.62	2.50	1.34
406	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Vehicle Type=Car	0.24	4.77	2.12	1.13

Table 162 – Association rules with vehicle characteristics as first antecedent and fatal crashes as consequent, Sweden.

ID Rule	Rules with vehicle characteristics as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
407	Vehicle Type=Truck	0.67	5.53	2.46	n.a.
408	Vehicle Type=Truck & Driver Gender=Male	0.58	8.80	3.91	1.59
409	Vehicle Type=Truck & Driver Gender=Male & Crash Location Detail=Road section	0.46	10.91	4.85	1.24
410	Vehicle Type=Truck & Driver Gender=Male & Crash Location Detail=Road section & N. Pedestrian involved=1	0.46	11.50	5.11	1.05
411	Vehicle Type=Truck & Driver Gender=Male & Crash Location=Not at intersection	0.49	10.80	4.80	1.23
412	Vehicle Type=Truck & Driver Gender=Male & Crash Location=Not at intersection & Pedestrian Gender=Male	0.28	11.45	5.09	1.06
413	Vehicle Type=Truck & Driver Gender=Male & Pedestrian Gender=Male	0.34	10.53	4.68	1.20
414	Vehicle Type=Truck & Driver Gender=Male & Pedestrian Gender=Male & Day of Week=Weekday	0.32	11.15	4.96	1.06
415	Vehicle Type=Truck & Crash Location Detail=Road section	0.53	8.05	3.58	1.46
416	Vehicle Type=Truck & Crash Location Detail=Road section & Driver Alcohol/Drug use=No	0.53	9.17	4.08	1.14
417	Vehicle Type=Truck & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & Pedestrian Gender=Male	0.32	9.84	4.37	1.07
418	Vehicle Type=Truck & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & N Pedestrian involved=1	0.53	9.63	4.28	1.05
419	Vehicle Type=Truck & Crash Location Detail=Road section & Driver Alcohol/Drug use=No & N Pedestrian involved=1 & Pedestrian Gender=Male	0.32	10.20	4.54	1.06
420	Vehicle Type=Truck & Crash Location Detail=Road section & Pedestrian Gender=Male	0.32	8.67	3.86	1.08
421	Vehicle Type=Truck & Pedestrian Gender=Male	0.40	6.95	3.09	1.26
422	Vehicle Type=Truck & Pedestrian Gender=Male & Crash Location=Not at intersection	0.34	8.16	3.63	1.18
423	Vehicle Type=Truck & Pedestrian Gender=Male & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.34	9.14	4.07	1.12
424	Vehicle Type=Truck & Pedestrian Gender=Male & Crash Location=Not at intersection & Day of Week=Weekday	0.32	8.82	3.92	1.08
425	Vehicle Type=Truck & Pedestrian Gender=Male & Crash Location=Not at intersection & Crash Location Detail=Road section	0.32	8.67	3.86	1.06
426	Vehicle Type=Truck & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.40	8.00	3.56	1.15
427	Vehicle Type=Truck & Pedestrian Gender=Male & Day of Week=Weekday	0.38	7.50	3.33	1.08
428	Vehicle Type=Truck & Pedestrian Gender=Male & Day of Week=Weekday & Crash Location Detail=Road section	0.30	9.30	4.14	1.24
429	Vehicle Type=Truck & Pedestrian Gender=Male & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.38	8.63	3.84	1.15
430	Vehicle Type=Truck & Crash Location=Not at intersection	0.56	6.89	3.06	1.25
431	Vehicle Type=Truck & Crash Location=Not at intersection & Driver Alcohol/Drug use=No	0.56	7.75	3.45	1.12
432	Vehicle Type=Truck & Crash Location=Not at intersection & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.53	9.17	4.08	1.18
433	Vehicle Type=Truck & Crash Location=Not at intersection & Driver Alcohol/Drug use=No & Crash Location Detail=Road section & Pedestrian Gender=Male	0.32	9.84	4.37	1.07
434	Vehicle Type=Truck & Crash Location=Not at intersection & Driver Alcohol/Drug use=No & Crash Location Detail=Road section & N Pedestrian involved=1	0.53	9.63	4.28	1.05
435	Vehicle Type=Truck & Driver Alcohol/Drug use=No	0.67	6.38	2.84	1.16

Table 163 – Association rules with driver characteristics as first antecedent and fatal crashes as consequent, Sweden.

ID Rule	Rules with driver characteristics as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
436	Driver Age=25-34	0.45	5.43	2.41	n.a.
437	Driver Age=25-34 & Pedestrian Gender=Male	0.22	6.54	2.91	1.21
438	Driver Age=25-34 & Pedestrian Gender=Male & N. vehicle involved=1	0.22	6.95	3.09	1.06
439	Driver Age=25-34 & Crash Location Detail=Road section	0.32	6.34	2.82	1.17
440	Driver Age=25-34 & Driver Gender=Male	0.38	6.16	2.74	1.14
441	Driver Age=25-34 & Driver Gender=Male & Crash Location Detail=Road section	0.27	7.10	3.16	1.15
442	Driver Age=25-34 & Driver Gender=Male & Day of Week=Weekday	0.33	6.78	3.02	1.10
443	Driver Age=25-34 & Driver Gender=Male & Pedestrian Gender=Female	0.20	6.60	2.93	1.07
444	Driver Age=55-64	0.38	4.36	1.94	n.a.
445	Driver Age=55-64 & Pedestrian Gender=Male	0.22	6.60	2.94	1.51
446	Driver Age=55-64 & Pedestrian Gender=Male & Crash Location=Not at intersection	0.21	8.26	3.67	1.25
447	Driver Age=55-64 & Crash Location Detail=Road section	0.31	5.69	2.53	1.30
448	Driver Age=55-64 & Driver Gender=Male	0.34	5.20	2.31	1.19
449	Driver Age=55-64 & Driver Gender=Male & Crash Location=Not at intersection	0.30	6.31	2.80	1.21
450	Driver Age=55-64 & Crash Location=Not at intersection	0.33	5.07	2.26	1.16
451	Driver Gender=Male	1.62	4.16	1.85	n.a.
452	Driver Gender=Male & Pedestrian Age=65-74	0.24	6.71	2.98	1.61
453	Driver Gender=Male & Lighting=Darkness	0.60	5.54	2.46	1.33
454	Driver Gender=Male & Lighting=Darkness & Pedestrian Gender=Male	0.40	7.58	3.37	1.37
455	Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section	0.48	6.97	3.10	1.26
456	Driver Gender=Male & Lighting=Darkness & Crash Location Detail=Road section & Pedestrian Gender=Male	0.35	9.43	4.19	1.35
457	Driver Gender=Male & Lighting=Darkness & Crash Location=Not at intersection	0.50	6.69	2.97	1.21
458	Driver Gender=Male & Lighting=Darkness & Crash Location=Not at intersection & Pedestrian Gender=Male	0.36	9.42	4.19	1.41
459	Driver Gender=Male & Lighting=Darkness & Day of Week=Weekday	0.49	6.13	2.72	1.11
460	Driver Gender=Male & Lighting=Darkness & Day of Week=Weekday & Pedestrian Gender=Male	0.32	8.50	3.78	1.39
461	Driver Gender=Male & Pedestrian Gender=Male	0.93	5.53	2.46	1.33
462	Driver Gender=Male & Pedestrian Gender=Male & Crash Location Detail=Road section	0.80	6.87	3.06	1.24
463	Driver Gender=Male & Pedestrian Gender=Male & Crash Location=Not at intersection	0.82	6.52	2.90	1.18
464	Driver Gender=Male & Crash Location Detail=Road section	1.32	5.40	2.40	1.30
465	Driver Gender=Male & Crash Location Detail=Road section & Driver Age=55-64	0.28	6.65	2.96	1.23
466	Driver Gender=Male & Crash Location Detail=Road section & Driver Age=55-64 & Pavement=Dry	0.28	7.14	3.18	1.07
467	Driver Gender=Male & Crash Location Detail=Road section & Driver Age=35-44	0.21	6.02	2.68	1.12
468	Driver Gender=Male & Crash Location=Not at intersection	1.36	4.78	2.13	1.15
469	Driver Gender=Male & Crash Location=Not at intersection & Driver Age=35-44	0.22	5.51	2.45	1.15
470	Driver Gender=Male & Crash Location=Not at intersection & Driver Age=35-44 & Crash Location Detail=Road section	0.21	6.02	2.68	1.09
471	Driver Gender=Male & Driver Age=35-44	0.24	4.40	1.96	1.06

472	Driver Age=45-54	0.34	3.74	1.66	n.a.
473	Driver Age=45-54 & Crash Location Detail=Road section	0.28	5.15	2.29	1.38
474	Driver Age=45-54 & Pedestrian Gender=Male	0.20	5.14	2.28	1.37
475	Driver Age=45-54 & Crash Location=Not at intersection	0.28	4.23	1.88	1.13
476	Driver Age=0-24	0.27	3.47	1.54	n.a.
477	Driver Age=0-24 & Pedestrian Gender=Male	0.20	5.94	2.64	1.71
478	Driver Age=0-24 & Crash Location Detail=Road section	0.23	5.06	2.25	1.46
479	Driver Age=0-24 & Crash Location=Not at intersection	0.23	4.05	1.80	1.17
480	Driver Age=35-44	0.27	3.47	1.54	n.a.
481	Driver Age=35-44 & Crash Location Detail=Road section	0.22	4.71	2.09	1.36
482	Driver Age=35-44 & Crash Location=Not at intersection	0.23	4.23	1.88	1.22

Rules with serious injury as consequent

Table 164 – Association rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent, Sweden.

ID Rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecedents	S %	C %	Lift	LIC
1	Pedestrian Alcohol/Drug use=Yes	0.27	25.00	5.53	n.a.
2	Pedestrian Alcohol/Drug use=Yes & Day of Week=Weekday	0.20	33.93	7.51	1.36
3	Pedestrian Alcohol/Drug use=Yes & Pedestrian Gender=Male	0.22	29.58	6.54	1.18
4	Pedestrian Alcohol/Drug use=Yes & Crash Location Detail=Road section	0.20	26.76	5.92	1.07
5	Pedestrian Alcohol/Drug use=Yes & Area=Urban	0.22	26.58	5.88	1.06
6	Pedestrian Alcohol/Drug use=Yes & N. Pedestrian involved=1	0.25	26.37	5.84	1.05
7	Pedestrian Alcohol/Drug use=Yes & N. Pedestrian involved=1 & Crash Location Detail=Road section	0.20	29.23	6.47	1.11
8	Pedestrian Alcohol/Drug use=Yes & N. Pedestrian involved=1 & Area=Urban	0.21	28.17	6.23	1.07
9	Pedestrian Age≥75	1.12	9.96	2.20	n.a.
10	Pedestrian Age≥75 & Lighting=Darkness	0.23	13.66	3.02	1.37
11	Pedestrian Age≥75 & Lighting=Darkness & Day of Week=Weekday	0.20	14.84	3.28	1.09
12	Pedestrian Age≥75 & Speed Limit=50	0.46	13.27	2.94	1.33
13	Pedestrian Age≥75 & Speed Limit=50 & Crash Location Detail=Road section	0.29	15.00	3.32	1.13
14	Pedestrian Age≥75 & Speed Limit=50 & Day of Week=Weekday	0.43	14.96	3.31	1.13
15	Pedestrian Age≥75 & Speed Limit=50 & Day of Week=Weekday & Crash Location Detail=Road section	0.27	16.56	3.66	1.11
16	Pedestrian Age≥75 & Speed Limit=50 & Day of Week=Weekday & Vehicle Type=Car	0.33	15.90	3.52	1.06
17	Pedestrian Age≥75 & Crash Location=At intersection	0.32	12.77	2.82	1.28
18	Pedestrian Age≥75 & Driver Gender=Female	0.25	12.63	2.79	1.27
19	Pedestrian Age≥75 & Driver Gender=Female & Pavement=Dry	0.25	14.12	3.12	1.12
20	Pedestrian Age≥75 & Speed Limit=40	0.27	11.16	2.47	1.12
21	Pedestrian Age≥75 & Speed Limit=40 & Day of Week=Weekday	0.23	12.02	2.66	1.08
22	Pedestrian Age≥75 & Speed Limit=40 & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.23	12.94	2.86	1.08
23	Pedestrian Age≥75 & Speed Limit=40 & Day of Week=Weekday & Area=Urban	0.22	12.80	2.83	1.07
24	Pedestrian Age≥75 & Speed Limit=40 & Area=Urban	0.25	12.00	2.66	1.08
25	Pedestrian Age≥75 & Speed Limit=40 & Area=Urban & Driver Alcohol/Drug use=No	0.25	12.70	2.81	1.06
26	Pedestrian Age≥75 & Speed Limit=40 & Lighting=Daylight	0.22	11.86	2.63	1.06
27	Pedestrian Age≥75 & Speed Limit=40 & Lighting=Daylight & Driver Alcohol/Drug use=No	0.22	12.96	2.87	1.09
28	Pedestrian Age≥75 & Speed Limit=40 & Road Type=Urban Municipal	0.22	11.73	2.60	1.05
29	Pedestrian Age≥75 & Speed Limit=40 & Road Type=Urban Municipal & Driver Alcohol/Drug use=No	0.22	12.50	2.77	1.07
30	Pedestrian Age≥75 & Pedestrian Gender=Male	0.43	11.14	2.47	1.12
31	Pedestrian Age≥75 & Pedestrian Gender=Male & Driver Gender=Male	0.23	12.87	2.85	1.15
32	Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban	0.37	12.03	2.66	1.08
33	Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban & Driver Gender=Male	0.20	13.67	3.02	1.14

ID Rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecedents	S %	C %	Lift	LIC
34	Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban & Vehicle Type=Car	0.31	13.00	2.88	1.08
35	Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban & Vehicle Type=Car & Lighting=Daylight	0.24	14.20	3.14	1.09
36	Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban & Vehicle Type=Car & Driver Alcohol/Drug use=No	0.31	13.74	3.04	1.06
37	Pedestrian Age≥75 & Pedestrian Gender=Male & Area=Urban & Driver Alcohol/Drug use=No	0.37	12.64	2.80	1.05
38	Pedestrian Age≥75 & Pedestrian Gender=Male & Driver Alcohol/Drug use=No	0.43	11.95	2.64	1.07
39	Pedestrian Age≥75 & Crash Location Detail=Intersection	0.24	10.90	2.41	1.09
40	Pedestrian Age≥75 & Driver Gender=Male	0.53	10.59	2.34	1.06
41	Pedestrian Age≥75 & Driver Gender=Male & Speed Limit=50	0.21	12.99	2.87	1.23
42	Pedestrian Age≥75 & Driver Gender=Male & Road Type=Urban Municipal	0.35	11.58	2.56	1.09
43	Pedestrian Age≥75 & Driver Gender=Male & Road Type=Urban Municipal & Crash Location=Not at intersection	0.24	12.78	2.83	1.10
44	Pedestrian Age≥75 & Driver Gender=Male & Crash Location=Not at intersection	0.41	11.21	2.48	1.06
45	Pedestrian Age≥75 & Driver Gender=Male & Crash Location=Not at intersection & Vehicle Type=Car	0.32	12.15	2.69	1.08
46	Pedestrian Age≥75 & Driver Gender=Male & Crash Location=Not at intersection & Vehicle Type=Car & Day of Week=Weekday	0.28	12.94	2.86	1.07
47	Pedestrian Age≥75 & Road Type=Urban Municipal	0.69	10.59	2.34	1.06
48	Pedestrian Age=65-74	0.72	8.05	1.78	n.a.
49	Pedestrian Age=65-74 & Driver Gender=Female	0.21	15.15	3.35	1.88
50	Pedestrian Age=65-74 & Speed Limit=40	0.22	12.07	2.67	1.50
51	Pedestrian Age=65-74 & Speed Limit=40 & Day of Week=Weekday	0.20	13.67	3.02	1.13
52	Pedestrian Age=65-74 & Speed Limit=40 & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.20	14.84	3.28	1.09
53	Pedestrian Age=65-74 & Road Type=Urban Municipal	0.48	9.26	2.05	1.15
54	Pedestrian Age=65-74 & Road Type=Urban Municipal & Day of Week=Weekday	0.41	9.97	2.21	1.08
55	Pedestrian Age=65-74 & Road Type=Urban Municipal & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.41	10.86	2.40	1.09
56	Pedestrian Age=65-74 & Road Type=Urban Municipal & Day of Week=Weekday & Pedestrian Gender=Female	0.27	10.55	2.33	1.06
57	Pedestrian Age=65-74 & Speed Limit=50	0.25	9.23	2.04	1.15
58	Pedestrian Age=65-74 & Speed Limit=50 & Day of Week=Weekday	0.23	10.38	2.30	1.12
59	Pedestrian Age=65-74 & Speed Limit=50 & Day of Week=Weekday & Driver Alcohol/Drug use=No	0.23	11.11	2.46	1.07
60	Pedestrian Age=65-74 & Speed Limit=50 & Day of Week=Weekday & Driver Alcohol/Drug use=No & N Pedestrian involved=1	0.23	11.70	2.59	1.05
61	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No	0.71	8.80	1.95	1.09
62	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Driver Gender=Female	0.21	15.27	3.38	1.73
63	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Speed Limit=40	0.22	13.29	2.94	1.51
64	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Road Type=Urban Municipal	0.48	10.14	2.24	1.15
65	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Speed Limit=50	0.25	9.88	2.19	1.12
66	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Crash Location Detail=Road section	0.45	9.84	2.18	1.12
67	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Crash Location Detail=Road section & Pedestrian Gender=Female	0.27	10.37	2.30	1.05
68	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Day of Week=Weekday	0.63	9.47	2.10	1.08
69	Pedestrian Age=65-74 & Driver Alcohol/Drug use=No & Day of Week=Weekday & Area=Urban	0.54	10.08	2.23	1.06
70	Pedestrian Age=65-74 & Day of Week=Weekday	0.64	8.77	1.94	1.09

ID Rule	Rules with pedestrian characteristics as first antecedent and serious injury crashes as consequent Antecedents	S %	C %	Lift	LIC
71	Pedestrian Age=65-74 & Crash Location Detail=Road section	0.45	8.77	1.94	1.09
72	Pedestrian Age=65-74 & Crash Location Detail=Road section & Road Type=Urban Municipal	0.28	10.57	2.34	1.21
73	Pedestrian Age=65-74 & Crash Location Detail=Road section & Road Type=Urban Municipal & Driver Alcohol/Drug use=No	0.28	11.66	2.58	1.10
74	Pedestrian Age=65-74 & Area=Urban	0.62	8.45	1.87	1.05
75	Pedestrian Age=65-74 & Area=Urban & Day of Week=Weekday	0.55	9.40	2.08	1.11
76	Pedestrian Age=65-74 & Area=Urban & Day of Week=Weekday & Pedestrian Gender=Female	0.35	9.94	2.20	1.06

Table 165 – Association rules with driver characteristics as first antecedent and serious injury crashes as consequent, Sweden.

ID Rule	Rules with driver characteristics as first antecedent and serious injury crashes as consequent Antecedents	S %	C %	Lift	LIC
77	N. vehicle involved=2	0.25	9.45	2.09	n.a.
78	Driver Age≥75	0.34	7.13	1.58	n.a.
79	Driver Age≥75 & Day of Week=Weekday	0.33	8.07	1.79	1.13
80	Driver Age≥75 & Day of Week=Weekday & Pedestrian Gender=Female	0.22	9.42	2.08	1.17
81	Driver Age≥75 & Pedestrian Gender=Female	0.22	7.89	1.75	1.11
82	Driver Age≥75 & Crash Location Detail=Road section	0.22	7.55	1.67	1.06
83	Driver Age≥75 & Crash Location Detail=Road section & Day of Week=Weekday	0.22	9.05	2.00	1.20
84	Driver Age=35-44	0.52	6.80	1.50	n.a.
85	Driver Age=35-44 & Lighting=Daylight	0.42	8.64	1.91	1.27
86	Driver Age=35-44 & Lighting=Daylight & Pedestrian Gender=Female	0.23	9.87	2.18	1.14
87	Driver Age=35-44 & Lighting=Daylight & Pedestrian Gender=Female & N. vehicle involved=1	0.23	10.48	2.32	1.06
88	Driver Age=35-44 & Pedestrian Gender=Female	0.28	7.45	1.65	1.10
89	Driver Age=35-44 & Pedestrian Gender=Female & Road Type=Urban Municipal	0.21	8.40	1.86	1.13
90	Driver Age=35-44 & Pedestrian Gender=Female & N. Pedestrian involved=1	0.28	7.88	1.74	1.06
91	Driver Age=35-44 & Pedestrian Gender=Male	0.23	7.14	1.58	1.05
92	Driver Age=35-44 & Pedestrian Gender=Male & Day of Week=Weekday	0.21	7.78	1.72	1.09
93	Driver Age=45-54	0.62	6.78	1.50	n.a.
94	Driver Age=45-54 & Lighting=Darkness	0.20	7.98	1.77	1.18

Table 166 – Association rules with roadway characteristics as first antecedent and serious injury crashes as consequent, Sweden.

ID Rule	Rules with roadway characteristics as first antecedent and serious injury crashes as consequent Antecedents	S %	C %	Lift	LIC
95	Road Type=Urban National	0.33	7.43	1.64	n.a.
96	Road Type=Urban National & Vehicle Type=Car	0.30	8.75	1.94	1.18
97	Road Type=Urban National & Vehicle Type=Car & Driver Alcohol/Drug use=No	0.30	10.18	2.25	1.16
98	Road Type=Urban National & Vehicle Type=Car & Driver Alcohol/Drug use=No & Day of Week=Weekday	0.25	10.81	2.39	1.06
99	Road Type=Urban National & Vehicle Type=Car & Driver Alcohol/Drug use=No & N. Pedestrian involved=1	0.30	10.73	2.37	1.05
100	Road Type=Urban National & Vehicle Type=Car & Day of Week=Weekday	0.25	9.64	2.13	1.10
101	Road Type=Urban National & Driver Alcohol/Drug use=No	0.33	8.49	1.88	1.14
102	Road Type=Urban National & Driver Alcohol/Drug use=No & Day of Week=Weekday	0.29	9.25	2.05	1.09
103	Road Type=Urban National & Driver Alcohol/Drug use=No & N. Pedestrian involved=1	0.33	8.93	1.98	1.05
104	Road Type=Urban National & Day of Week=Weekday	0.29	8.31	1.84	1.12
105	Road Type=Rural National	0.25	6.84	1.51	n.a.

Artificial neural network

Table 167 – Artificial Neural Network parameter estimates, Sweden.

Predictor		Predicted				
		Hidden Layer 1				
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)
Input Layer	(Bias)	-.128	-.539	.263	-.269	.127
	Day of Week=Weekday	.219	-.003	-.099	-.296	-.293
	Day of Week=Weekend	.023	-.477	-.478	-.365	-.178
	Area=Rural	.100	-.080	-.326	.519	.195
	Area=Urban	.156	.360	.199	.095	.155
	Road Type=Motorway	-.259	-.150	.106	.347	.036
	Road Type=Rural Individual	-.217	-.153	-.282	.340	.121
	Road Type=Rural Municipal	-.626	.336	-.377	-.057	.062
	Road Type=Rural National	-.219	.266	.275	.169	-.088
	Road Type=Urban Individual	.396	.420	-.384	-.124	-.222
	Road Type=Urban Municipal	-.700	-.334	-.493	.083	.189
	Road Type=Urban National	.362	.098	-.384	-.060	-.003
	Crash Location=At intersection	.272	.006	-.253	-.006	.264
	Crash Location=Not at intersection	-.026	.045	.277	-.537	.398
	Crash Location Detail=Interchange	.202	.284	-.079	-.175	.034
	Crash Location Detail=Intersection	.220	-.122	-.245	-.209	-.070
	Crash Location Detail=Pedestrian/bicyclepath	.158	.268	.256	-.256	-.079
	Crash Location Detail=Road section	.460	.116	-.090	.244	-.009
	Crash Location Detail=Roundabout	.143	-.497	.178	-.216	-.243
	Crash Location Detail=Separate parking space	-.099	.385	.203	-.506	.189
	Pavement=Dry	.369	.313	.429	-.307	.195
	Pavement=Slippery	.239	-.180	.098	.437	-.384
	Pavement=Snowy/icy	.319	-.224	.438	-.570	-.097
	Pavement=Unevenness	.137	-.178	.138	-.003	-.382
	Pavement=Wet	.303	.517	-.111	.059	.328
	Lighting=Darkness	.185	-.177	-.205	-.242	.417
	Lighting=Dawn/dusk	-.434	.505	.399	.172	-.283
	Lighting=Daylight	.252	.049	.422	-.080	.242

Predictor		Predicted				
		Hidden Layer 1				
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)
	Speed Limit=<30	-.297	-.426	.356	.364	.008
	Speed Limit=40	-.044	.373	.239	.108	.109
	Speed Limit=50,	.254	-.160	-.378	.074	.036
	Speed Limit=60+	.396	-.331	-.299	.187	-.547
	N. vehicle involved=1	.142	-.547	.235	-.172	.359
	N. vehicle involved=2	-.187	.262	.110	-.051	-.223
	Vehicle Type=Bike	-.401	-.388	-.531	.059	.104
	Vehicle Type=Car	-.110	.280	-.072	.159	-.102
	Vehicle Type=PTW	-.156	.261	.361	.347	-.383
	Vehicle Type=Truck	-.140	.141	-.341	.325	-.447
	Driver Gender=Female	-.330	.050	.369	.165	-.451
	Driver Gender=Male	-.510	-.011	-.347	.268	-.334
	Driver Age≤75	-.036	.103	-.460	-.196	-.005
	Driver Age=0-24	-.062	-.317	.112	-.250	-.393
	Driver Age=25-34	-.264	.141	-.199	.064	-.546
	Driver Age=35-44	-.241	-.444	.032	.276	.165
	Driver Age=45-54	-.027	-.086	.351	-.305	-.390
	Driver Age=55-64	.189	-.314	.085	-.139	.327
	Driver Age=65-74	.085	.413	-.188	-.277	.357
	Driver Alcohol/Drug use=No	-.393	-.013	.198	.391	-.037
	Vehicle N Trailers=0,	.417	-.457	.109	.303	-.547
	Vehicle N Trailers=1,	.531	.431	.271	-.250	.334
	N Pedestrian involved=1,	-.220	.249	.170	.273	-.071
	N Pedestrian involved=2,	-.151	-.470	.184	.300	-.156
	N Pedestrian involved=2+	-.029	-.453	-.069	.322	.254
	Pedestrian Gender=Female	.334	.291	-.031	-.179	-.070
	Pedestrian Gender=Male	.048	-.313	-.068	-.379	-.311
	Pedestrian Age≤75	.445	.274	.404	.195	-.988
	Pedestrian Age=0-14	-.435	.450	-.109	-.067	-.113
	Pedestrian Age=15-24	-.068	.072	.294	.108	.395
	Pedestrian Age=25-34	.062	.173	-.491	-.097	.409
	Pedestrian Age=35-44	.516	-.489	.256	-.370	.256

Predictor		Predicted				
		Hidden Layer 1				
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)
	Pedestrian Age=45-54	-.091	.305	-.008	.089	.383
	Pedestrian Age=55-64	.269	-.149	.060	.480	.200
	Pedestrian Age=65-74	.379	.310	-.367	-.211	-.200
	Pedestrian Alcohol/Drug use=Yes	.437	.301	-.376	-.506	.319

Table 168 – Artificial Neural Network parameter estimates for the output layer, Sweden.

Predictor		Predicted		
		Output Layer		
		Crash Severity=Fatal	Crash Severity=Serious	Crash Severity=Slight
Hidden Layer 1	(Bias)	0.030	0.212	-0.390
	H(1:1)	0.790	-0.178	-0.343
	H(1:2)	-0.708	0.087	0.141
	H(1:3)	0.452	0.047	-0.241
	H(1:4)	1.039	-0.054	-0.182
	H(1:5)	-1.332	0.272	1.056

APPENDIX 3 ~ ITALY

Classification tree

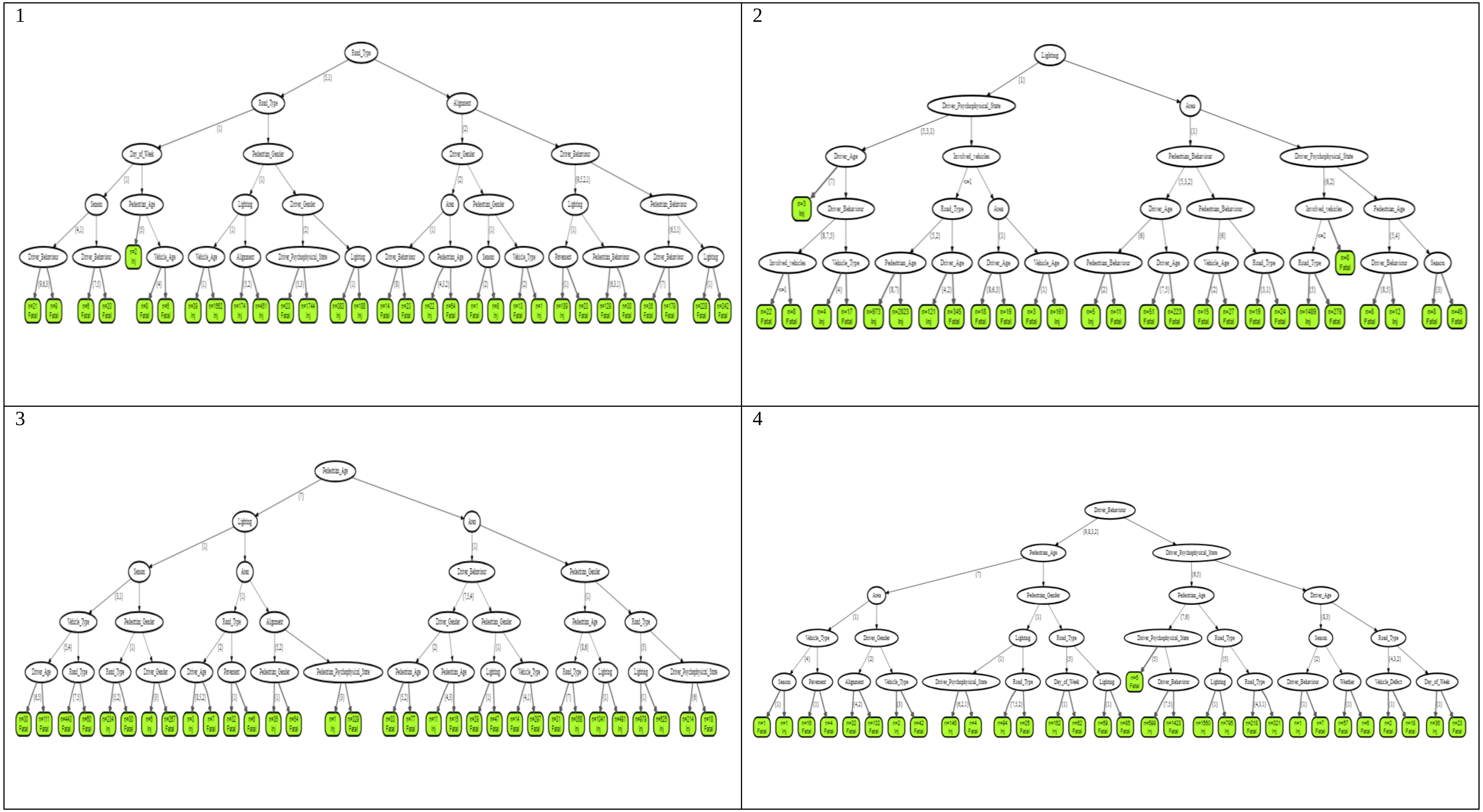
Table 169 – Tree in table format, Italy.

Node	Injury		Fatal		Total		Predicted Category	Parent Node
	N	%	N	%	N	%		
0	98,063	97.1	2,969	2.9	101,032	100.0	0	-
1	86,210	97.9	1,813	2.1	88,023	74.5	0	0
2	11,853	91.1	1,156	8.9	13,009	25.5	1	0
3	59,425	99.0	591	1.0	60,016	40.3	0	1
4	26,785	95.6	1,222	4.4	28,007	34.2	1	1
5	2,715	83.8	525	16.2	3,240	10.2	1	2
6	9,138	93.5	631	6.5	9,769	15.3	1	2
7	40,722	99.4	242	0.6	40,964	24.8	0	3
8	18,703	98.2	349	1.8	19,052	15.4	0	3
9	16,152	94.6	925	5.4	17,077	23.8	1	4
10	10,633	97.3	297	2.7	10,930	10.4	0	4
11	1,302	90.2	141	9.8	1,443	3.0	1	5
12	1,413	78.6	384	21.4	1,797	7.2	1	5
13	5,077	96.9	164	3.1	5,241	5.4	1	6
14	4,061	89.7	467	10.3	4,528	9.9	1	6
15	232	93.2	17	6.8	249	0.4	1	7
16	40,490	99.4	225	0.6	40,715	24.4	0	7
17	16,766	98.5	262	1.5	17,028	13.0	0	8
18	1,937	95.7	87	4.3	2,024	2.5	1	8
19	13,103	95.4	630	4.6	13,733	17.3	1	9
20	3,049	91.2	295	8.8	3,344	6.5	1	9
21	8,092	97.8	181	2.2	8,273	7.2	0	10
22	2,541	95.6	116	4.4	2,657	3.2	1	10
23	797	93.1	59	6.9	856	1.4	1	11
24	505	86.0	82	14.0	587	1.6	1	11
25	723	85.7	121	14.3	844	2.4	1	12
26	690	72.4	263	27.6	953	4.8	1	12
27	3,103	98.4	49	1.6	3,152	2.4	0	13
28	1,974	94.5	115	5.5	2,089	2.9	1	13
29	1,737	85.8	287	14.2	2,024	5.7	1	14
30	2,324	92.8	180	7.2	2,504	4.2	1	14

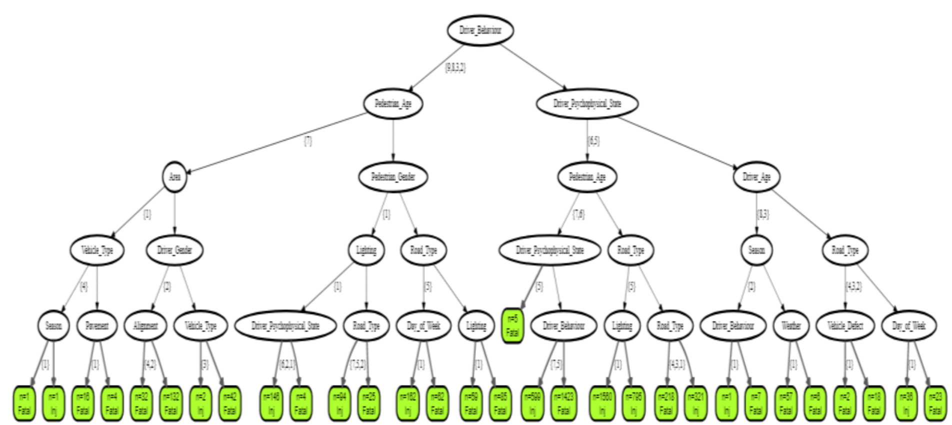
Table 170 – Posterior Classification Ratio (PCR) for all nodes, Italy.

Node	PCR		
	Injury	Fatal	Actual Predicted Class
0	1.00	1.00	-
1	0.88	0.61	Injury
2	0.12	0.39	Fatal
3	0.61	0.20	Injury
4	0.27	0.41	Fatal
5	0.03	0.18	Fatal
6	0.09	0.21	Fatal
7	0.42	0.08	Injury
8	0.19	0.12	Injury
9	0.16	0.31	Fatal
10	0.11	0.10	Injury
11	0.01	0.05	Fatal
12	0.01	0.13	Fatal
13	0.05	0.06	Fatal
14	0.04	0.16	Fatal
15	0.00	0.01	Fatal
16	0.41	0.08	Injury
17	0.17	0.09	Injury
18	0.02	0.03	Fatal
19	0.13	0.21	Fatal
20	0.03	0.10	Fatal
21	0.08	0.06	Injury
22	0.03	0.04	Fatal
23	0.01	0.02	Fatal
24	0.01	0.03	Fatal
25	0.01	0.04	Fatal
26	0.01	0.09	Fatal
27	0.03	0.02	Injury
28	0.02	0.04	Fatal
29	0.02	0.10	Fatal
30	0.02	0.06	Fatal

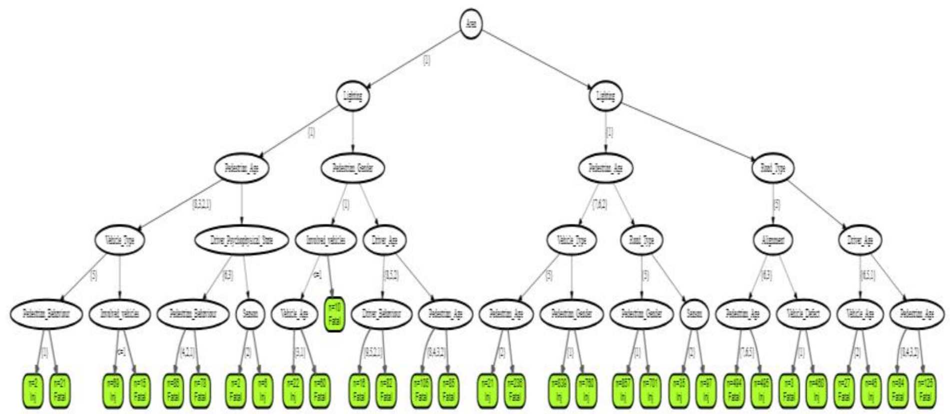
Random forest



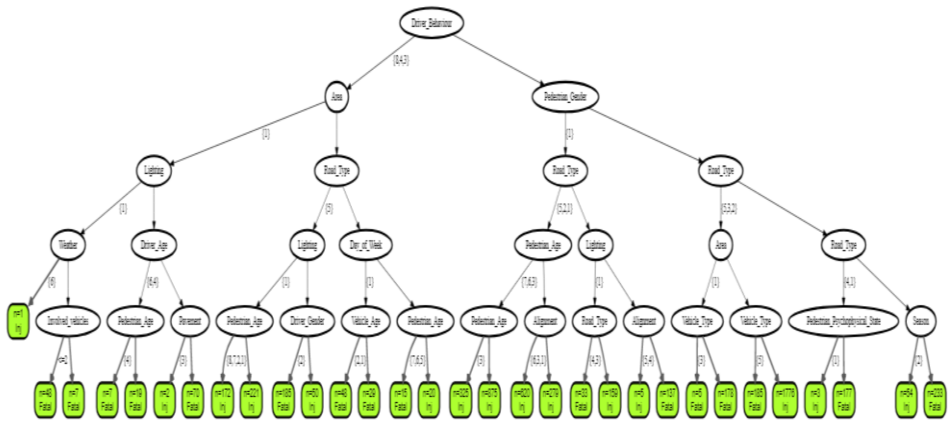
5



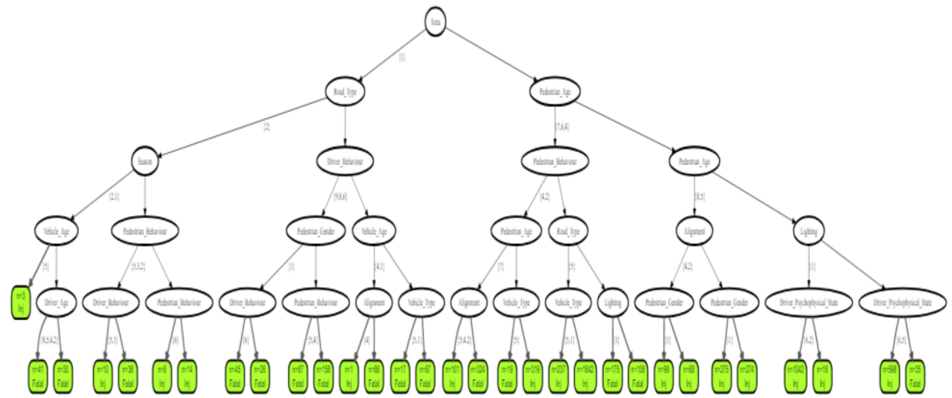
6



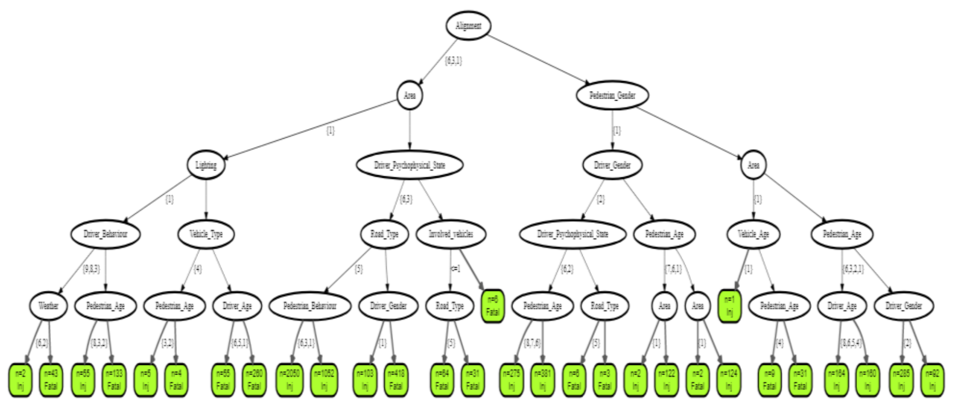
7



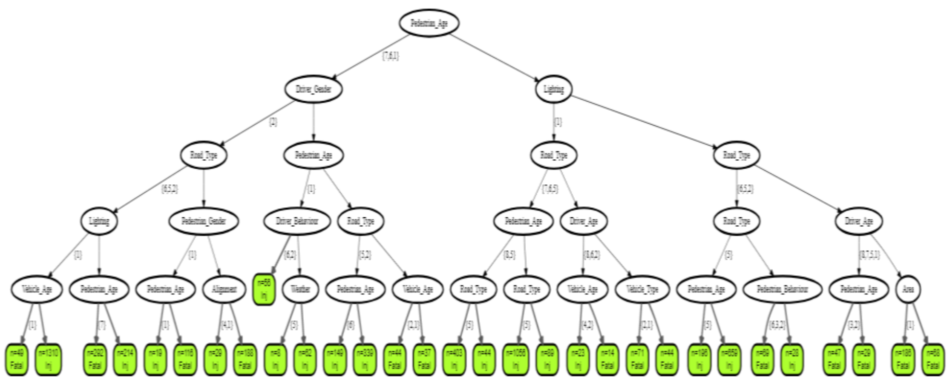
8



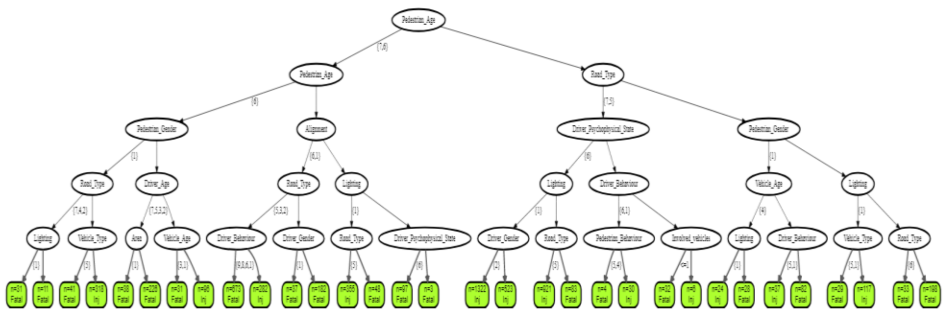
9



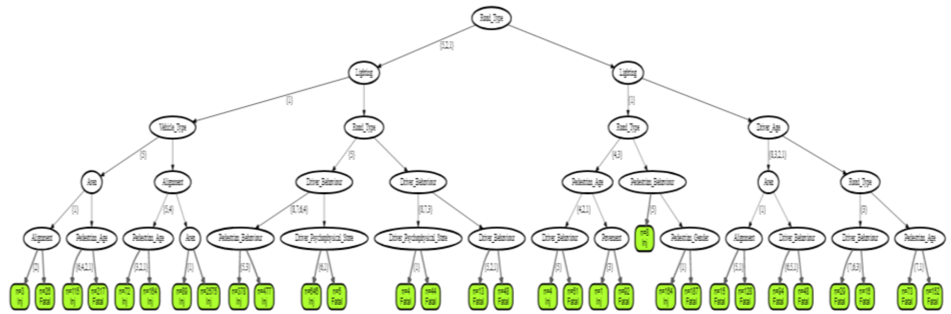
10



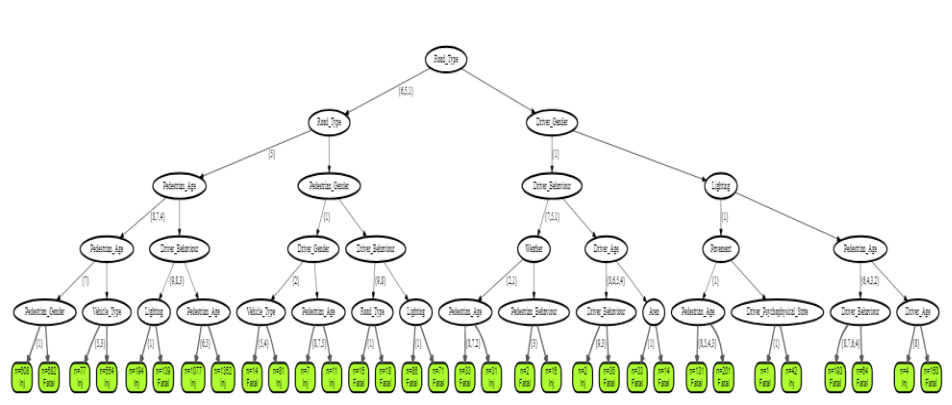
11



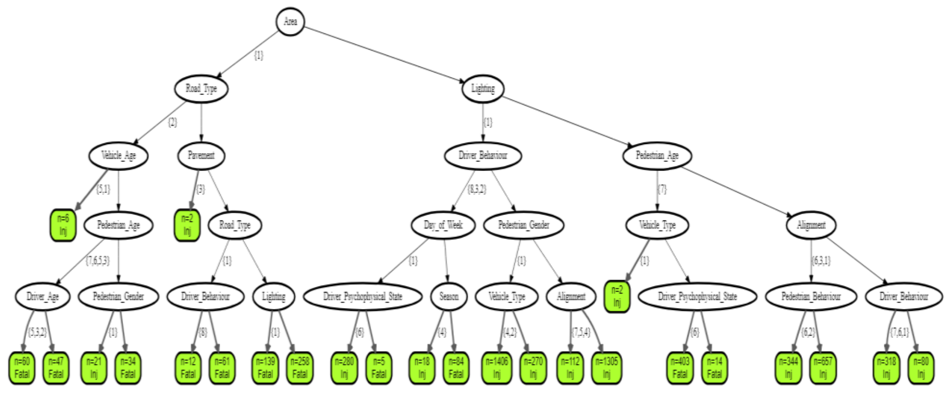
12



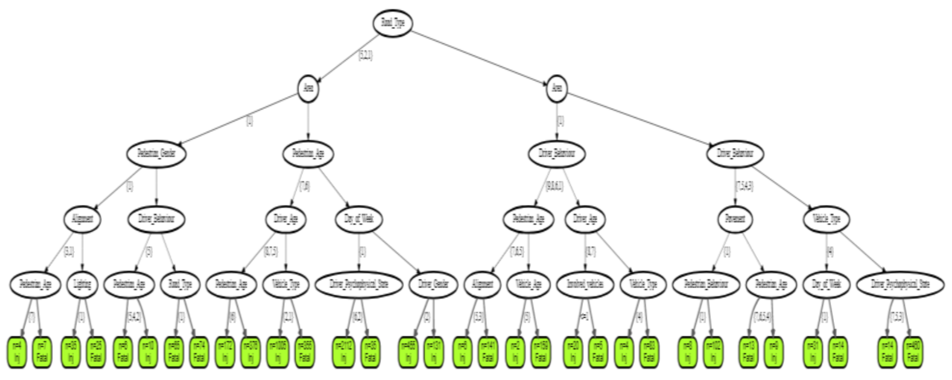
13



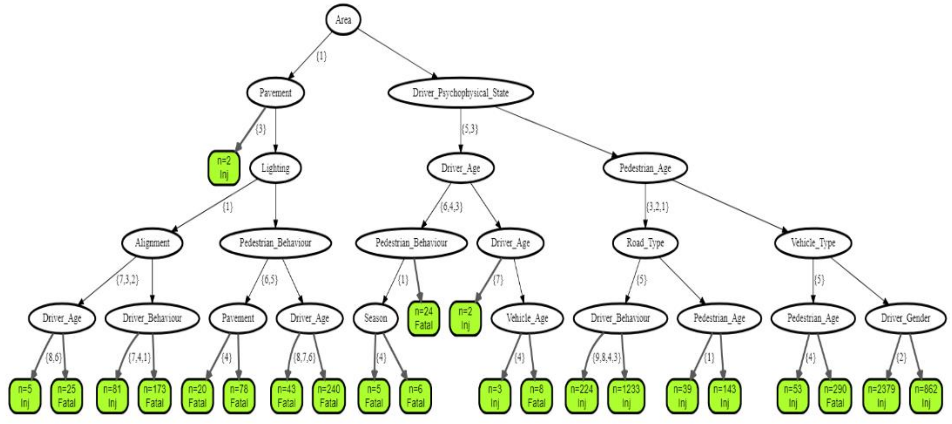
14


Area
Road_Type
Lighting
Vehicle_Age
Pavement
Driver_Behaviour
Pedestrian_Age
Pedestrian_Gender
Day_of_Week
Vehicle_Type
Alignment
Driver_Age
Driver_Psychophysical_State
Season
Pedestrian_Behaviour

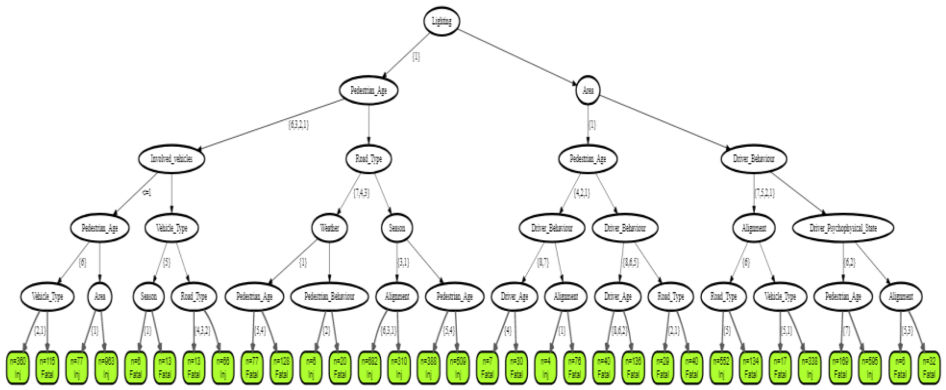
15



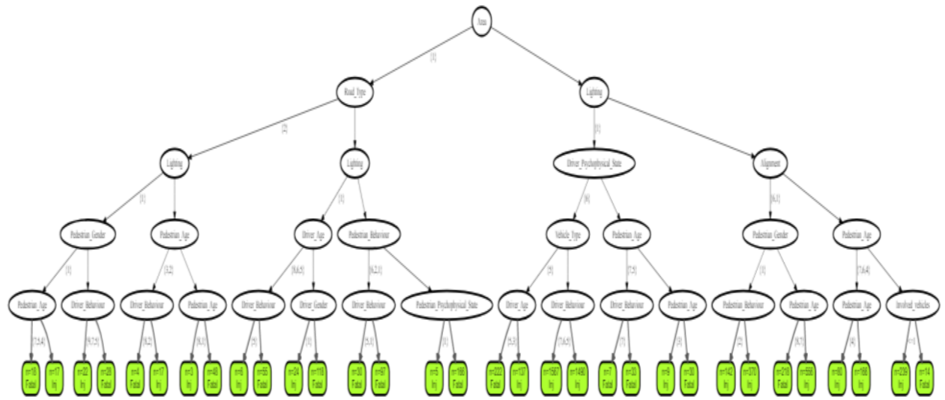
16


Area
Pavement
Driver_Psychophysical_State
Lighting
Driver_Age
Pedestrian_Age
Alignment
Pedestrian_Behaviour
Road_Type
Vehicle_Type
Driver_Behaviour
Season
Vehicle_Age
Driver_Gender

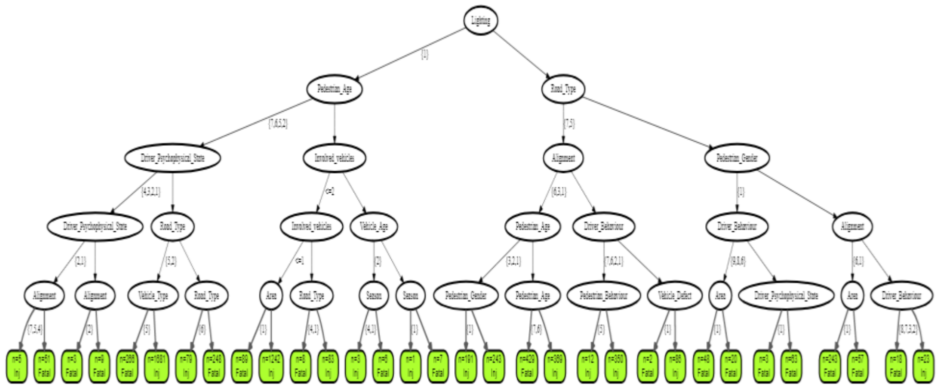
17



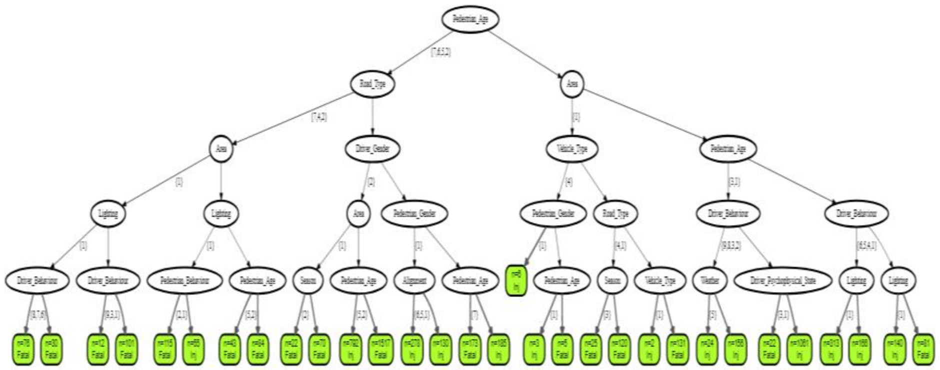
18



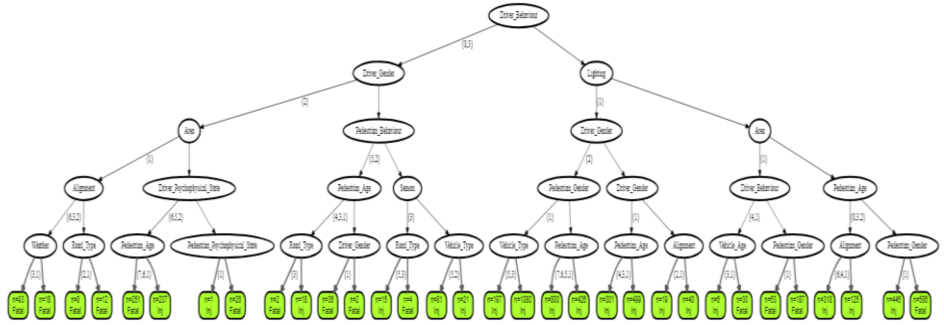
19



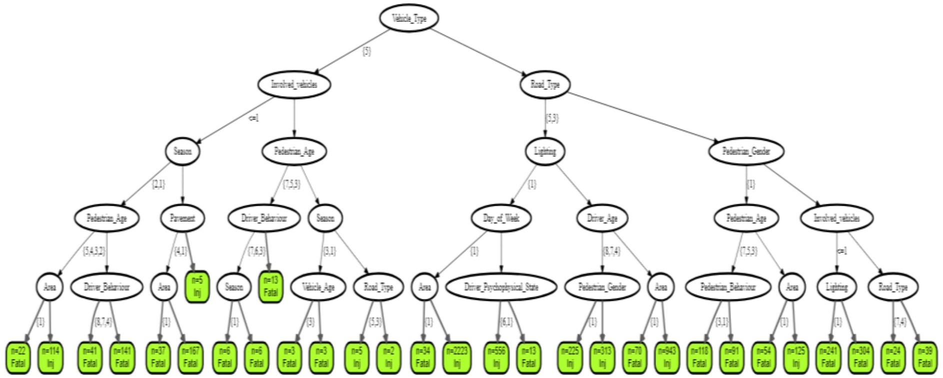
20



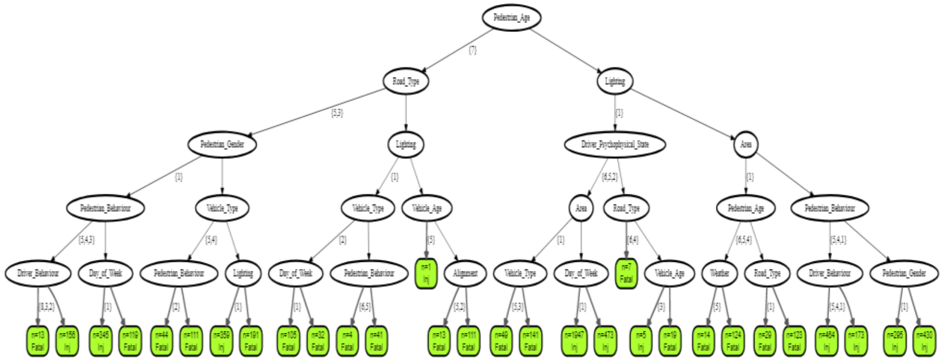
21



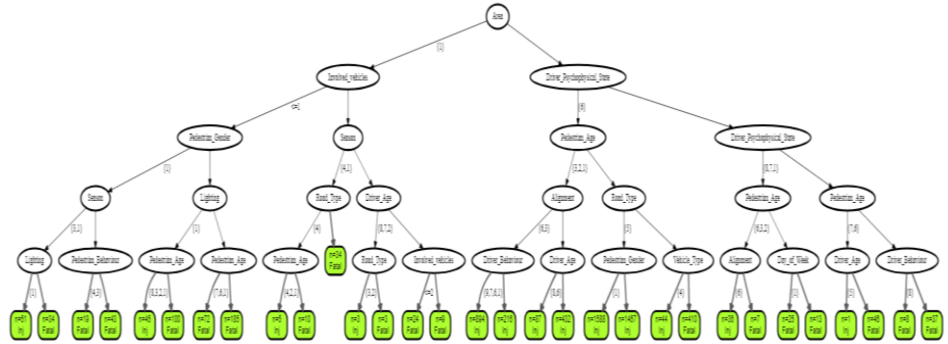
22



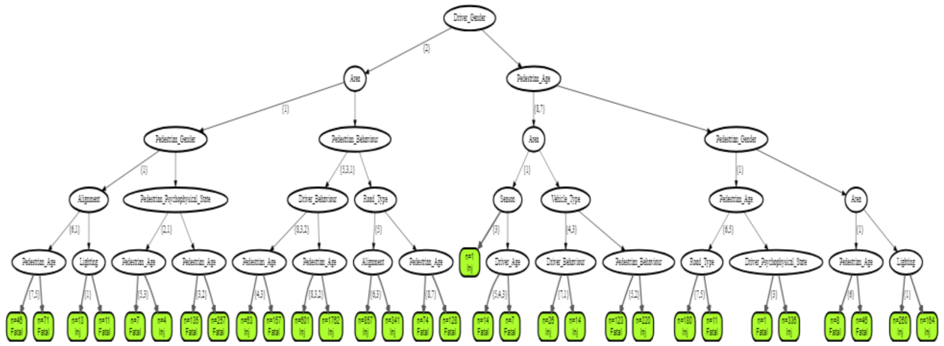
23



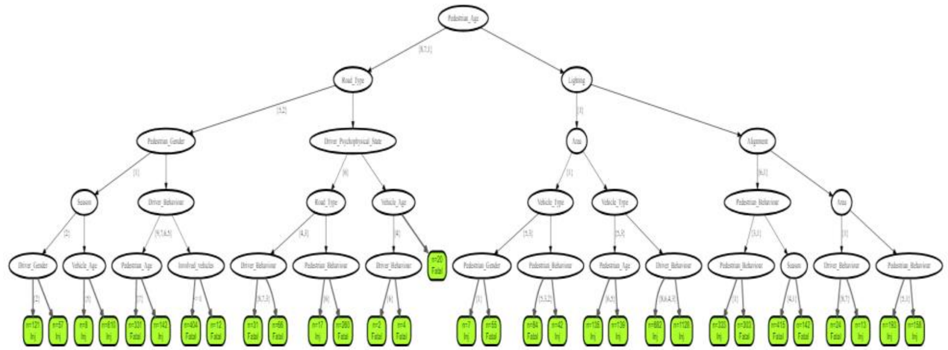
24



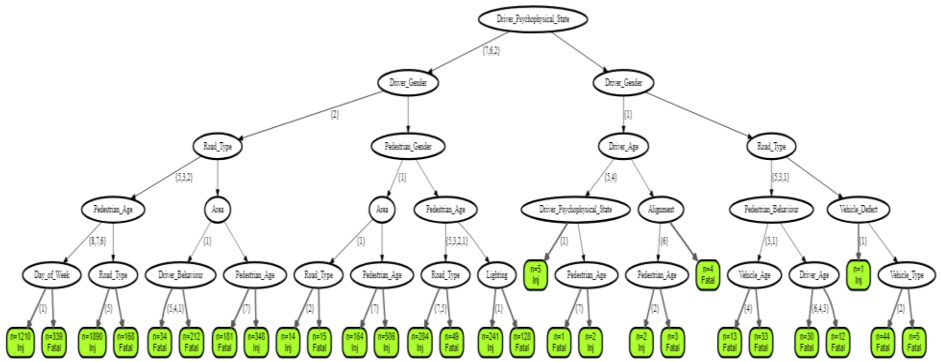
25



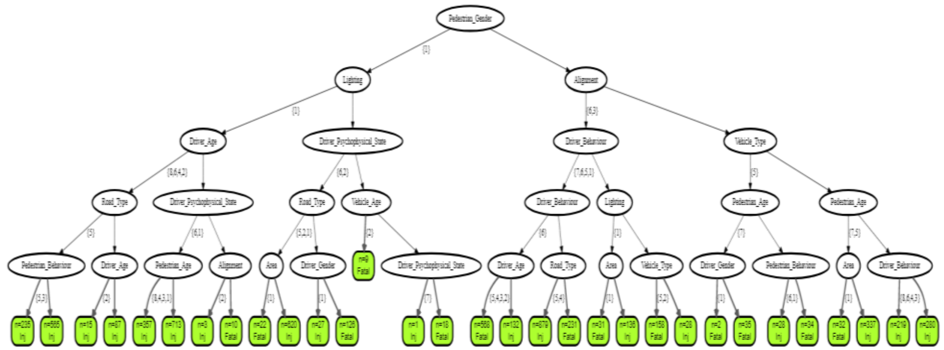
26



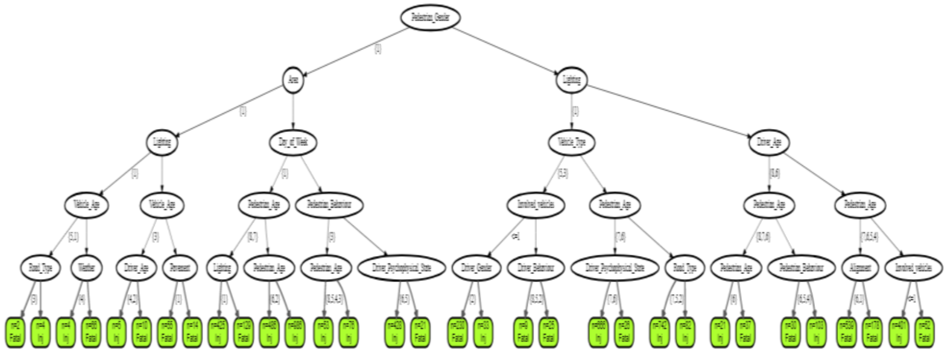
27



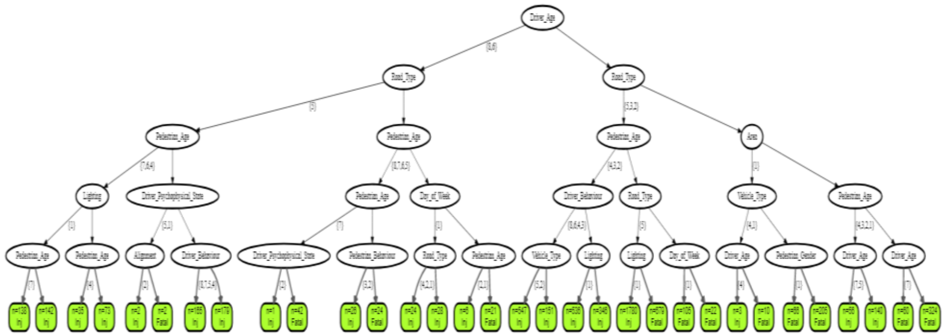
28



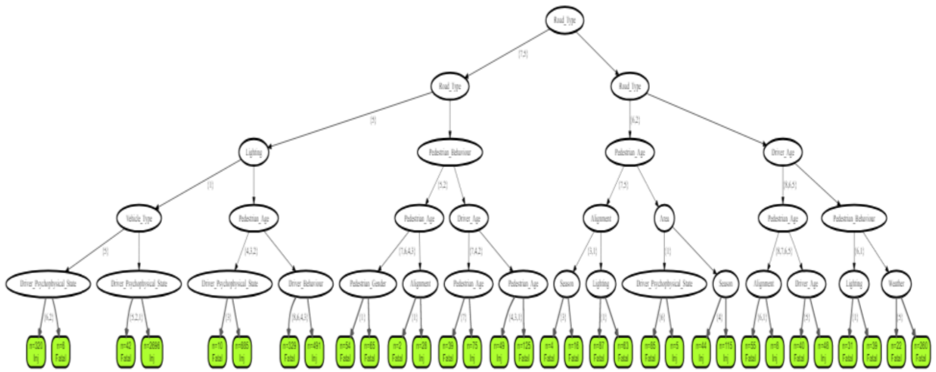
29



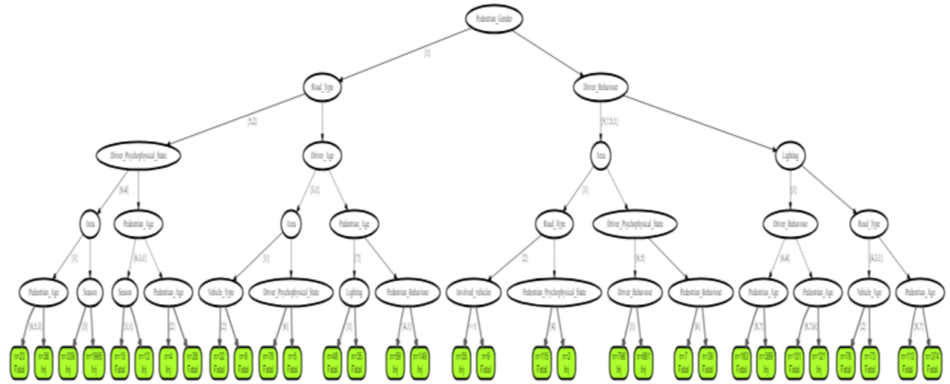
30



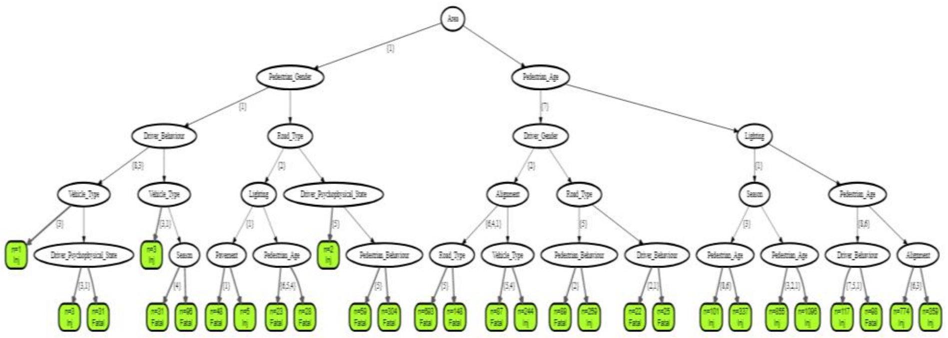
31



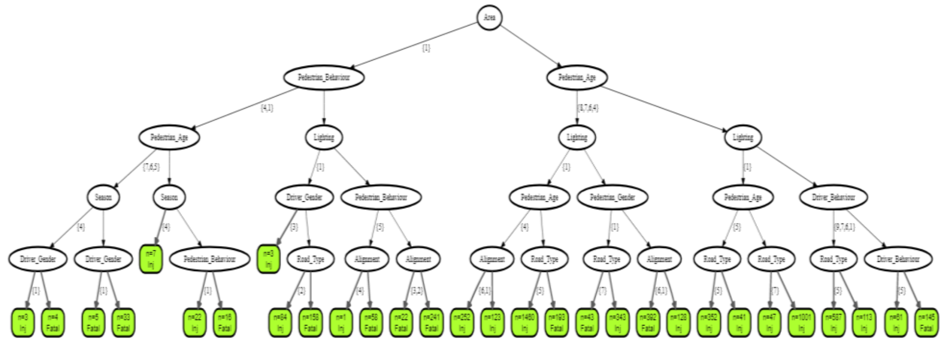
32



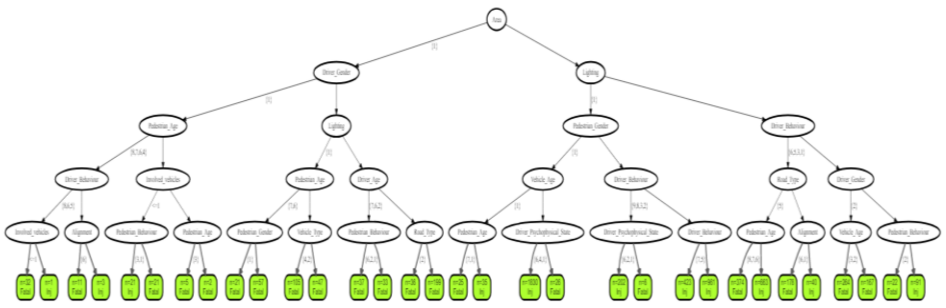
33



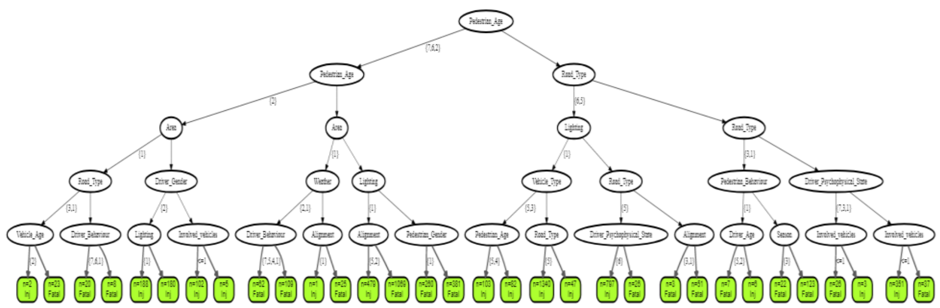
34



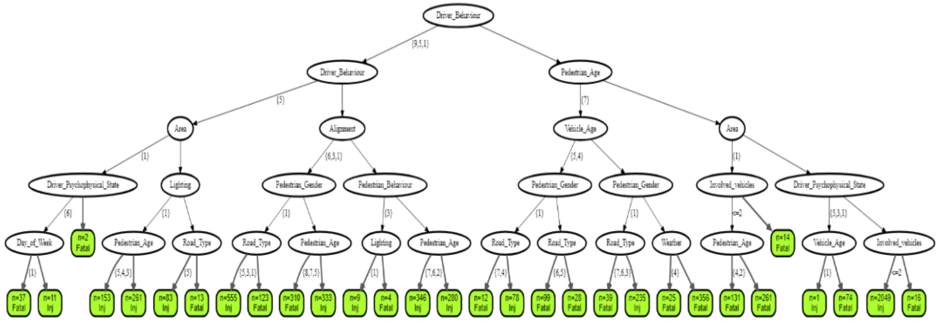
35



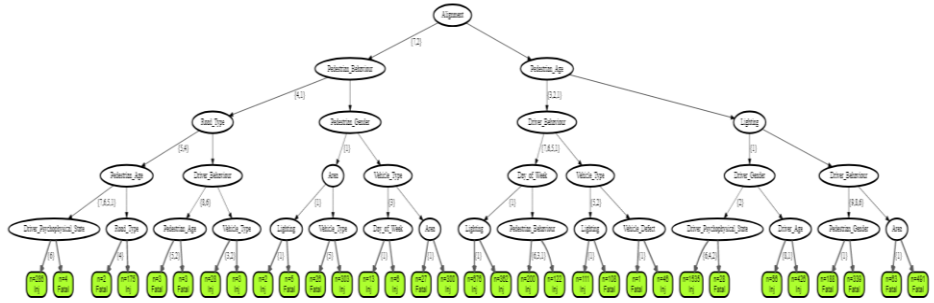
36



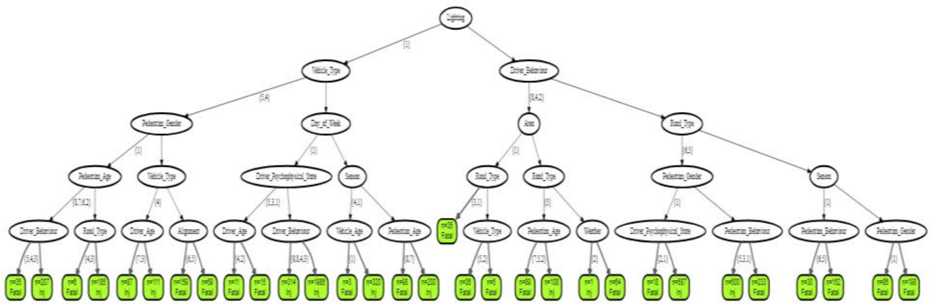
37



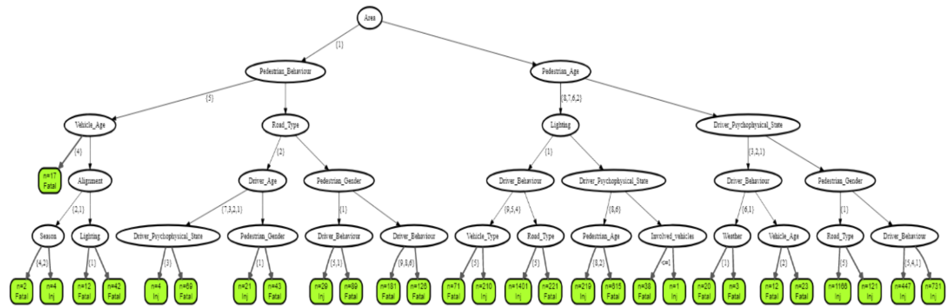
38



39



40



Association rules

Rules with fatal as consequent

Table 171 – Association rules with pedestrian characteristics as first antecedent and fatal crashes as consequent, Italy.

ID Rule	Rules with roadway characteristics as first antecedent and fatal crashes as consequent	S	C	Lift	LIC
	Antecedents	%	%		
1	Road Type=Rural national	17.32	16.53	5.62	n.a.
2	Road Type= Rural national & Lighting=Night	12.47	22.99	7.82	1.39
3	Road Type=Rural provincial	25.64	13.99	4.76	n.a.
4	Road Type= Rural provincial & Lighting=Night	16.23	18.87	6.42	1.35
5	Area=Rural	63.94	13.24	4.51	n.a.
6	Area=Rural & Driver behaviour=Speed	13.96	23.46	7.98	1.77
7	Area=Rural & Driver behaviour=Speed & Pavement=Dry	12.27	25.00	8.51	1.07
8	Area=Rural & Pedestrian Age≥ 75	13.36	20.61	7.01	1.56
9	Area=Rural & Pedestrian Age≥ 75 & Vehicle Type=Car	11.09	21.71	7.39	1.05
10	Area=Rural & Lighting=Night	40.68	19.98	6.80	1.51
11	Area=Rural & Season=Winter	17.22	14.09	4.79	1.06
12	Road Type=Rural municipal	11.98	7.39	2.51	n.a.
13	Road Type=Urban provincial	33.26	6.76	2.30	n.a.
14	Road Type=Urban provincial & Pedestrian Age≥ 75	16.83	14.82	5.04	2.19
15	Road Type=Urban provincial & Pedestrian Age≥ 75 & Alignment=Tangent	12.87	16.84	5.73	1.14
16	Road Type=Urban provincial & Pedestrian Age≥ 75 & Alignment=Tangent & Driver Gender=Male	10.69	18.40	6.26	1.09
17	Road Type=Urban provincial & Pedestrian Age≥ 75 & Driver Gender=Male	13.76	16.16	5.50	1.09
18	Road Type=Urban provincial & Pedestrian Age≥ 75 & Day of Week=Weekday	13.36	15.77	5.37	1.06
19	Road Type=Urban provincial & Pedestrian Age≥ 75 & Day of Week=Weekday & Alignment=Tangent	19.58	18.02	6.13	1.14
20	Road Type=Urban provincial & Pedestrian Age≥ 75 & Day of Week=Weekday & Driver Gender=Male	11.28	17.95	6.11	1.14
21	Road Type=Urban provincial & Lighting=Night	17.12	9.64	3.28	1.43
22	Road Type=Urban provincial & Season=Autumn	12.77	7.21	2.46	1.07
23	Road Type=Urban national	17.22	5.50	1.87	n.a.

Table 172 – Association rules with pedestrian characteristics as first antecedent and fatal crashes as consequent, Italy.

ID Rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent	S	C	Lift	LIC
	Antecedents	**%**	**%**		
24	Pedestrian Age≥ 75	1.30	6.75	2.30	n.a.
25	Pedestrian Age≥ 75 & Driver behaviour=Speed	19.70	12.05	4.10	1.79
26	Pedestrian Age≥ 75 & Driver behaviour=Speed & Pedestrian Gender=Male	12.77	15.54	5.29	1.29
27	Pedestrian Age≥ 75 & Driver behaviour=Speed & Pedestrian Gender=Male & Driver Gender=Male	10.89	17.32	5.89	1.11
28	Pedestrian Age≥ 75 & Driver behaviour=Speed & Driver Gender=Male	16.73	13.78	4.69	1.14
29	Pedestrian Age≥ 75 & Driver behaviour=Speed & Alignment=Tangent	13.96	13.49	4.59	1.12
30	Pedestrian Age≥ 75 & Driver behaviour=Speed & Alignment=Tangent & Driver Gender=Male	11.88	15.13	5.15	1.12
31	Pedestrian Age≥ 75 & Driver behaviour=Speed & Vehicle Age=0-10	10.29	13.47	4.58	1.12
32	Pedestrian Age≥ 75 & Driver behaviour=Speed & Vehicle Type=Car	15.54	13.05	4.44	1.08
33	Pedestrian Age≥ 75 & Driver behaviour=Speed & Vehicle Type=Car & Driver Gender=Male	12.87	15.10	5.14	1.16
34	Pedestrian Age≥ 75 & Lighting=Night	46.92	11.73	3.99	1.74
35	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	17.72	16.20	5.51	1.38
36	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Alignment=Tangent	14.45	19.49	6.63	1.20
37	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Pedestrian Gender=Male	12.17	19.46	6.62	1.20
38	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Vehicle Age=0-10	10.29	17.72	6.03	1.09
39	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Pavement=Dry	12.08	17.23	5.86	1.06
40	Pedestrian Age≥ 75 & Lighting=Night & Driver behaviour=Normal	15.04	14.84	5.04	1.26
41	Pedestrian Age≥ 75 & Lighting=Night & Driver behaviour=Normal & Pedestrian Gender=Male	10.29	17.48	5.95	1.18
42	Pedestrian Age≥ 75 & Lighting=Night & Driver behaviour=Normal & Alignment=Tangent	11.78	17.42	5.93	1.18
43	Pedestrian Age≥ 75 & Lighting=Night & Driver Age=45-54	11.58	14.41	4.90	1.23
44	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Gender=Male	30.98	14.31	4.87	1.22
45	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Gender=Male & Driver Age=25-44	10.49	17.04	5.80	1.19
46	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Gender=Male & Alignment=Tangent	23.66	16.47	5.61	1.15
47	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Gender=Male & Driver Gender=Male	25.54	15.28	5.20	1.07
48	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Gender=Male & Weather=Clear	21.78	15.21	5.18	1.06
49	Pedestrian Age≥ 75 & Lighting=Night & Pedestrian Gender=Male & Pavement=Dry	21.08	15.17	5.16	1.06
50	Pedestrian Age≥ 75 & Lighting=Night & Alignment=Tangent	36.23	13.96	4.75	1.19
51	Pedestrian Age≥ 75 & Lighting=Night & Alignment=Tangent & Driver Age=25-44	11.88	17.24	5.87	1.24
52	Pedestrian Age≥ 75 & Lighting=Night & Alignment=Tangent & Vehicle Age=0-10	19.80	15.89	5.41	1.14
53	Pedestrian Age≥ 75 & Lighting=Night & Alignment=Tangent & Vehicle Type=Car	29.99	14.79	5.03	1.06
54	Pedestrian Age≥ 75 & Lighting=Night & Alignment=Tangent & Driver Gender=Male	29.50	14.75	5.02	1.06
55	Pedestrian Age≥ 75 & Lighting=Night & Driver Age=25-44	15.04	13.68	4.66	1.17
56	Pedestrian Age≥ 75 & Lighting=Night & Driver Age=25-44 & Vehicle Type=Car	13.57	15.24	5.19	1.11
57	Pedestrian Age≥ 75 & Lighting=Night & Driver Age=25-44 & Driver Gender=Male	12.37	14.86	5.06	1.09
58	Pedestrian Age≥ 75 & Lighting=Night & Vehicle Age=0-10	25.34	12.96	4.41	1.10

ID Rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent	S	C	Lift	LIC
	Antecedents	%	%		
59	Pedestrian Age≥ 75 & Lighting=Night & Vehicle Age=0-10 & Driver Gender=Male	21.28	14.19	4.83	1.10
60	Pedestrian Age≥ 75 & Lighting=Night & Driver Gender=Male	38.40	12.73	4.33	1.09
61	Pedestrian Age≥ 75 & Lighting=Night & Day of Week=Weekend	12.37	12.35	4.20	1.05
62	Pedestrian Age≥ 75 & Lighting=Night & Day of Week=Weekend & Alignment=Tangent	10.19	15.47	5.26	1.25
63	Pedestrian Age≥ 75 & Lighting=Night & Day of Week=Weekend & Vehicle Type=Car	10.99	13.21	4.50	1.07
64	Pedestrian Age≥ 75 & Lighting=Night & Day of Week=Weekend & Driver Gender=Male	10.19	13.14	4.47	1.06
65	Pedestrian Age≥ 75 & Vehicle Type=Truck	17.72	10.37	3.53	1.54
66	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	41.67	8.97	3.05	1.33
67	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Pedestrian Gender=Male	26.03	11.40	3.88	1.27
68	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Day of Week=Weekend	10.19	10.60	3.61	1.18
69	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Alignment=Tangent	29.69	10.26	3.49	1.14
70	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Alignment=Tangent & Pedestrian Gender=Male	18.31	12.85	4.37	1.25
71	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Season=Autumn	15.64	9.78	3.33	1.09
72	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Driver Age=25-44	14.15	9.72	3.31	1.08
73	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Driver Gender=Male	32.96	9.54	3.25	1.06
74	Pedestrian Age≥ 75 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing & Driver Gender=Male & Pedestrian Gender=Male	21.08	12.07	4.11	1.27
75	Pedestrian Age≥ 75 & Pedestrian Gender=Male	78.19	8.71	2.96	1.29
76	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Driver behaviour=Normal	22.57	10.67	3.63	1.23
77	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Driver Age=25-44	26.13	9.88	3.36	1.14
78	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Driver Age=45-54	17.62	9.72	3.31	1.12
79	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Alignment=Tangent	52.95	9.71	3.30	1.12
80	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Season=Autumn	30.98	9.55	3.25	1.10
81	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Driver Gender=Male	63.54	9.50	3.23	1.09
82	Pedestrian Age≥ 75 & Pedestrian Gender=Male & Day of Week=Weekend	17.32	9.30	3.16	1.07
83	Pedestrian Age≥ 75 & Driver Age=18-24	11.28	8.32	2.83	1.23
84	Pedestrian Age≥ 75 & Driver behaviour=Normal	34.44	8.28	2.82	1.23
85	Pedestrian Age≥ 75 & Driver behaviour=Normal & Alignment=Tangent	24.15	9.69	3.30	1.17
86	Pedestrian Age≥ 75 & Driver behaviour=Normal & Alignment=Tangent & Pedestrian Gender=Male	14.85	11.65	3.96	1.20
87	Pedestrian Age≥ 75 & Driver behaviour=Normal & Alignment=Tangent & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	15.74	10.96	3.73	1.13
88	Pedestrian Age≥ 75 & Driver behaviour=Normal & Driver Age=25-44	12.87	9.37	3.19	1.13
89	Pedestrian Age≥ 75 & Driver behaviour=Normal & Driver Age=25-44 & Driver Gender=Male	10.29	10.42	3.55	1.11
90	Pedestrian Age≥ 75 & Driver behaviour=Normal & Season=Autumn	12.97	8.79	2.99	1.06
91	Pedestrian Age≥ 75 & Day of Week=Weekend	31.48	7.82	2.67	1.16
92	Pedestrian Age≥ 75 & Day of Week=Weekend & Alignment=Tangent	22.86	9.45	3.22	1.21
93	Pedestrian Age≥ 75 & Day of Week=Weekend & Alignment=Tangent & Vehicle Age=0-10	12.57	11.23	3.82	1.19
94	Pedestrian Age≥ 75 & Day of Week=Weekend & Alignment=Tangent & Pedestrian Gender=Male	12.47	10.91	3.71	1.15
95	Pedestrian Age≥ 75 & Day of Week=Weekend & Alignment=Tangent & Driver Gender=Male	18.71	9.96	3.39	1.05

ID Rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
96	Pedestrian Age≥ 75 & Day of Week=Weekend & Vehicle Age=0-10	16.03	8.39	2.85	1.07
97	Pedestrian Age≥ 75 & Day of Week=Weekend & Vehicle Age=0-10 & Driver Gender=Male	12.87	8.87	3.02	1.06
98	Pedestrian Age≥ 75 & Day of Week=Weekend & Driver Gender=Male	25.64	8.35	2.84	1.07
99	Pedestrian Age≥ 75 & Day of Week=Weekend & Driver Gender=Male & Pedestrian Gender=Male	13.66	9.64	3.28	1.15
100	Pedestrian Age≥ 75 & Alignment=Tangent	89.48	7.73	2.63	1.15
101	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=25-44	29.69	9.06	3.08	1.17
102	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=25-44 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	10.99	12.36	4.21	1.36
103	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=25-44 & Pedestrian Gender=Male	17.92	11.24	3.83	1.24
104	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=25-44 & Driver Gender=Male	24.35	10.23	3.48	1.13
105	Pedestrian Age≥ 75 & Alignment=Tangent & Weather=Rainy	10.89	8.85	3.01	1.14
106	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Autumn	36.33	8.70	2.96	1.13
107	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Autumn & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	11.68	11.78	4.74	1.35
108	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Autumn & Pedestrian Gender=Male	21.48	10.97	3.73	1.26
109	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Autumn & Vehicle Age=0-10	19.40	9.40	3.24	1.08
110	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Autumn & Driver behaviour=Disobeying Pedestrian Crossings	14.05	9.34	3.18	1.07
111	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Autumn & Driver Gender=Male	28.80	9.29	3.16	1.07
112	Pedestrian Age≥ 75 & Alignment=Tangent & Pavement=Wet	15.34	8.56	2.91	1.11
113	Pedestrian Age≥ 75 & Alignment=Tangent & Pavement=Wet & Driver Gender=Male	12.77	9.31	3.17	1.09
114	Pedestrian Age≥ 75 & Alignment=Tangent & Driver behaviour= Disobeying Pedestrian Crossings	33.55	8.48	2.89	1.10
115	Pedestrian Age≥ 75 & Alignment=Tangent & Driver behaviour= Disobeying Pedestrian Crossings & Pedestrian Gender=Male	19.20	14.66	3.42	1.18
116	Pedestrian Age≥ 75 & Alignment=Tangent & Driver behaviour= Disobeying Pedestrian Crossings & Driver Gender=Male	26.53	9.09	3.09	1.07
117	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=45-54	19.10	8.45	2.87	1.09
118	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=45-54 & Pedestrian Gender=Male	11.78	10.95	3.73	1.30
119	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Age=45-54 & Driver Gender=Male	15.64	9.27	3.16	1.10
120	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Winter	24.84	8.41	2.86	1.09
121	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Winter & Pedestrian Gender=Male	15.44	10.59	3.60	1.26
122	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Winter & Vehicle Age=0-10	13.76	9.55	3.25	1.13
123	Pedestrian Age≥ 75 & Alignment=Tangent & Season=Winter & Driver Gender=Male	20.19	9.14	3.11	1.09
124	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Gender=Male	72.55	8.39	2.86	1.09
125	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Gender=Male & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	23.95	10.83	3.68	1.29
126	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Gender=Male & Pedestrian Gender=Male	43.25	10.45	3.56	1.25
127	Pedestrian Age≥ 75 & Alignment=Tangent & Driver Gender=Male & Vehicle Age=0-10	37.81	8.92	3.04	1.06
128	Pedestrian Age≥ 75 & Alignment=Tangent & Vehicle Age=0-10	46.42	8.25	2.81	1.07
129	Pedestrian Age≥ 75 & Alignment=Tangent & Vehicle Age=0-10 & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	15.84	10.58	3.60	1.28
130	Pedestrian Age≥ 75 & Alignment=Tangent & Vehicle Age=0-10 & Driver behaviour=Normal	13.36	10.42	3.55	1.26
131	Pedestrian Age≥ 75 & Alignment=Tangent & Vehicle Age=0-10 & Pedestrian Gender=Male	25.83	9.74	3.31	1.18
132	Pedestrian Age≥ 75 & Driver Age=25-44	42.36	7.63	2.60	1.13

ID Rule	Rules with pedestrian characteristics as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
133	Pedestrian Age≥ 75 & Driver Age=25-44 & Driver Gender=Male	34.05	8.71	2.96	1.14
134	Pedestrian Age≥ 75 & Driver Age=25-44 & Driver Gender=Male & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	11.28	10.87	3.70	1.25
135	Pedestrian Age≥ 75 & Driver Age=25-44 & Driver Gender=Male & Pedestrian Gender=Male	20.19	10.51	3.58	1.21
136	Pedestrian Age≥ 75 & Driver Age=25-44 & Season=Winter	11.78	8.23	2.84	1.08
137	Pedestrian Age≥ 75 & Vehicle Type=PTW	13.36	7.46	2.54	1.11
138	Pedestrian Age≥ 75 & Driver Gender=Male	1.05	7.44	2.53	1.10
139	Pedestrian Age≥ 75 & Driver Gender=Male & Weather=Rainy	12.27	8.31	2.83	1.12
140	Pedestrian Age≥ 75 & Driver Gender=Male & Pavement=Wet	18.21	8.30	2.82	1.12
141	Pedestrian Age≥ 75 & Driver Gender=Male & Pavement=Wet & Pedestrian Gender=Male	11.48	10.50	3.57	1.26
142	Pedestrian Age≥ 75 & Driver Gender=Male & Driver Age=45-54	22.57	8.25	2.81	1.11
143	Pedestrian Age≥ 75 & Driver Gender=Male & Driver Age=45-54 & Pedestrian Gender=Male	15.34	11.75	4.00	1.43
144	Pedestrian Age≥ 75 & Driver Gender=Male & Season=Autumn	40.48	7.97	2.71	1.07
145	Pedestrian Age≥ 75 & Driver Gender=Male & Season=Autumn & Pedestrian Behaviour=No Crossing on Pedestrian Crossing	12.27	10.31	3.51	1.29
146	Pedestrian Age≥ 75 & Driver Gender=Male & Season=Autumn & Pedestrian Gender=Male	24.74	10.16	3.46	1.28
147	Pedestrian Age≥ 75 & Season=Autumn	51.27	7.37	2.51	1.09
148	Pedestrian Age≥ 75 & Season=Autumn & Driver Age=45-54	11.68	8.43	2.87	1.15
149	Pedestrian Age≥ 75 & Season=Autumn & Vehicle Age=0-10	27.71	7.94	2.70	1.08
150	Pedestrian Age≥ 75 & Season=Autumn & Vehicle Age=0-10 & Lighting=Night	12.17	13.17	4.48	1.66
151	Pedestrian Age≥ 75 & Season=Autumn & Vehicle Age=0-10 & Pedestrian Gender=Male	15.84	9.66	3.29	1.22
152	Pedestrian Age≥ 75 & Season=Autumn & Vehicle Age=0-10 & Driver Gender=Male	22.07	8.61	2.93	1.08
153	Pedestrian Age≥ 75 & Driver Age=45-54	27.81	7.32	2.49	1.09
154	Pedestrian Age≥ 75 & Driver Age=45-54 & Vehicle Age=0-10	16.13	7.78	2.65	1.06
155	Pedestrian Age≥ 75 & Driver Age=45-54 & Vehicle Age=0-10 & Driver Gender=Male	12.87	8.67	2.95	1.11
156	Pedestrian Age≥ 75 & Pavement=Wet	21.28	7.25	2.47	1.07
157	Pedestrian Age≥ 75 & Pavement=Wet & Vehicle Age=0-10	11.58	7.63	2.60	1.05
158	Pedestrian Age≥ 75 & Weather=Rainy	14.25	7.23	2.46	1.07

Table 173 – Association rules with vehicle characteristics as first antecedent and fatal crashes as consequent, Italy.

ID Rule	Rules with vehicle characteristics as first antecedent and fatal crashes as consequent Antecedents	S %	C %	Lift	LIC
159	Vehicle Type=Truck	41.37	5.97	2.03	n.a.
160	Vehicle Type=Truck & Lighting=Night	11.09	8.66	2.95	1.45
161	Vehicle Type=Truck & Season=Winter	10.99	6.47	2.20	1.08
162	Vehicle Type=Truck & Season=Autumn	15.74	6.35	2.16	1.06
163	Vehicle Type=Truck & Season=Autumn & Alignment=Tangent	10.49	6.73	2.29	1.06
164	Vehicle Type=Truck & Alignment=Tangent	27.52	6.31	2.15	1.06

Table 174 – Association rules with driver characteristics as first antecedent and fatal crashes as consequent, Italy.

ID Rule	Rules with driver characteristics as first antecedent and fatal crashes as consequent	S	C	Lift	LIC
	Antecedents	**%**	**%**		
165	Driver behaviour=Speed	53.94	5.79	1.97	n.a.
166	Driver behaviour=Speed & Lighting=Night	30.49	9.83	3.35	1.70
167	Driver behaviour=Speed & Lighting=Night & Vehicle Age=0-10	15.84	11.89	4.05	1.21
168	Driver behaviour=Speed & Lighting=Night & Vehicle Age=0-10 & Pavement=Dry	12.67	14.08	4.79	1.18
169	Driver behaviour=Speed & Lighting=Night & Vehicle Age=0-10 & Weather=Clear	12.47	13.49	4.59	1.13
170	Driver behaviour=Speed & Lighting=Night & Vehicle Age=0-10 & Vehicle Type=Car	12.57	13.16	4.48	1.11
171	Driver behaviour=Speed & Lighting=Night & Vehicle Age=0-10 & Alignment=Tangent	11.28	12.58	4.28	1.06
172	Driver behaviour=Speed & Lighting=Night & Day of Week=Weekend	19.58	11.68	3.98	1.19
173	Driver behaviour=Speed & Lighting=Night & Pavement=Dry	24.35	11.33	3.85	1.15
174	Driver behaviour=Speed & Lighting=Night & Pavement=Dry & Vehicle Type=Car	19.40	12.89	4.39	1.14
175	Driver behaviour=Speed & Lighting=Night & Vehicle Type=Car	24.84	11.21	3.81	1.14
176	Driver behaviour=Speed & Lighting=Night & Vehicle Type=Car & Weather=Clear	19.10	12.52	4.26	1.12
177	Driver behaviour=Speed & Lighting=Night & Vehicle Type=Car & Alignment=Tangent	18.61	12.50	4.25	1.12
178	Driver behaviour=Speed & Lighting=Night & Weather=Clear	23.85	10.97	3.73	1.12
179	Driver behaviour=Speed & Lighting=Night & Alignment=Tangent	22.37	10.47	3.56	1.06
180	Driver behaviour=Speed & Day of Week=Weekend	16.73	8.01	2.73	1.38
181	Driver behaviour=Speed & Day of Week=Weekend & Vehicle Type=Car	13.36	8.88	3.02	1.11
182	Driver behaviour=Speed & Day of Week=Weekend & Vehicle Type=Car & Pavement=Dry	11.88	9.59	3.26	1.08
183	Driver behaviour=Speed & Day of Week=Weekend & Vehicle Type=Car & Weather=Clear	11.48	9.53	3.24	1.07
184	Driver behaviour=Speed & Day of Week=Weekend & Pavement=Dry	15.04	8.66	2.95	1.08
185	Driver behaviour=Speed & Day of Week=Weekend & Weather=Clear	14.35	8.48	2.89	1.06
186	Driver behaviour=Speed & Vehicle Type=Car	41.57	6.32	2.15	1.09
187	Driver behaviour=Speed & Vehicle Type=Car & Alignment=Tangent	29.99	7.21	2.45	1.14
188	Driver behaviour=Speed & Alignment=Tangent	38.60	6.31	2.15	1.09
189	Driver behaviour=Speed & Alignment=Tangent & Season=Autumn	14.55	6.78	2.31	1.07
190	Driver behaviour=Speed & Alignment=Tangent & Season=Autumn & Vehicle Type=Car	11.28	7.65	2.60	1.13

Table 175 – Association rules with environmental characteristics as first antecedent and fatal crashes as consequent, Italy.

ID Rule	Rules with environmental characteristics as first antecedent and fatal crashes as consequent	S	C	Lift	LIC
	Antecedents	**%**	**%**		
191	Lighting=Night	1.43	4.79	1.63	n.a.
192	Lighting=Night & Season=Summer	14.55	6.65	2.26	1.39
193	Lighting=Night & Season=Springr	17.42	5.89	2.57	1.23
194	Lighting=Night & Day of Week=Weekend	43.06	5.60	1.91	1.17

Artificial neural network

Table 176 – Artificial Neural Network parameter estimates, Italy.

Predictor		Predicted													
		Hidden Layer 1													
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)	H(1:14)
Input Layer	(Bias)	.290	.216	.207	-.102	-.015	.378	.145	-.036	-.106	-.295	-.562	-.170	.129	.361
	Day of Week=Weekday	-.536	.116	-.229	-.374	-.179	-.423	.324	.155	.329	-.269	-.255	.103	-.521	.299
	Day of Week=Weekend	-.254	.086	.151	.418	.286	-.427	.445	.165	.069	.355	-.292	-.402	-.042	.338
	Season=Autumn	-.075	-.442	-.280	.234	.134	-.077	-.217	.124	-.039	-.207	-.080	.412	.425	-.029
	Season=Spring	.174	-.262	-.440	.198	.008	.151	-.489	.115	.304	.473	.110	-.444	-.147	-.463
	Season=Summer	-.516	-.423	.155	-.407	.349	-.314	.513	-.262	.088	.494	-.256	.337	.218	-.384
	Season=Winter	-.374	.367	-.214	.497	.327	-.243	.456	.090	.254	-.040	.329	.343	.183	.120
	Lighting=Day	-.481	.154	.099	.275	.174	.372	.005	-.458	.431	.329	-.100	-.351	-.142	.248
	Lighting=Night	-.083	-.313	-.521	.267	-.506	-.074	.367	.251	-.241	.165	.210	.157	.089	.041
	Road Type=Motorway	.439	.039	-.321	.403	-.011	.455	-.083	-.136	.349	.017	-.357	-.202	.298	.358
	Road Type=Rural Municipal	-.382	.539	.090	-.082	-.294	.004	.056	-.031	-.419	-.146	.339	.162	.142	-.465
	Road Type=Rural national	-.231	.233	-.088	-.102	-.499	.421	-.368	.350	-.174	.324	.068	.530	.005	-.080
	Road Type=Rural provincial	.302	.081	.214	-.389	-.317	-.439	.216	-.008	.138	.184	.216	-.386	.483	.133
	Road Type=Urban Municipal	-.351	.555	.377	-.130	.418	-.358	.338	-.255	-.310	-.432	-.282	-.418	-.396	-.478
	Road Type=Urban national	.456	.281	-.502	.159	.399	.304	-.200	.003	-.136	.119	.354	-.275	.263	.330
	Road Type=Urban provincial	-.104	.168	-.545	.498	-.227	-.028	.383	-.244	.314	.451	-.416	.417	.162	.081
	Area=Rural	.482	-.297	-.567	.269	-.143	-.499	-.105	-.154	-.131	.545	.305	-.083	-.130	-.289
	Area=Urban	-.486	.209	.099	-.301	-.180	.189	.040	-.241	.221	-.561	-.244	-.076	-.281	.099
	Alignment=Curve	.207	-.535	-.326	-.063	-.182	.439	-.066	.326	.388	-.278	-.277	-.173	.490	-.441
	Alignment=No Segnalized Intersection	-.205	.133	-.413	-.121	.511	-.105	-.407	-.058	.115	.124	.163	.260	.476	-.229
	Alignment=Roundabout	.040	.079	.017	.059	-.036	-.174	.217	-.362	-.307	-.292	-.326	-.407	-.188	-.168
	Alignment=Segnalized Intersection	.460	.159	-.386	.411	.205	-.372	-.076	.329	-.502	.321	.385	.227	-.369	-.208
	Alignment=Tangent	.337	-.583	.067	-.117	-.022	-.384	-.055	.195	-.410	-.056	-.557	.002	.261	-.395
	Alignment=Tunnel	.129	.439	.413	-.030	.280	-.320	-.415	-.419	-.086	.367	.325	.482	-.042	-.028
	Pavement=Dry	.080	.317	.128	.339	.213	.325	-.114	-.053	-.201	.192	-.450	.168	.157	.301
	Pavement=Slippery	-.330	.434	.385	-.197	-.399	.440	.221	-.273	-.480	.183	.258	.068	.004	.151
	Pavement=Snowy/Frozen	.348	-.466	-.101	-.012	.106	.076	-.413	-.321	.070	.031	-.405	-.002	.491	-.214

Predictor	Predicted													
	Hidden Layer 1													
	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)	H(1:14)
Pavement=Wet	-.107	-.104	-.361	-.273	.262	.154	-.438	-.234	-.183	-.077	.143	.358	-.056	-.277
Weather=Clear	.376	.171	-.363	.251	.111	-.165	-.247	.502	-.050	.153	-.330	-.380	.044	.338
Weather=Fog	.460	.207	.263	-.355	-.227	.454	.393	.004	.214	.060	-.281	.409	-.042	-.316
Weather=High winds	-.448	-.051	-.297	.110	.415	.406	.160	.344	-.118	.170	.356	-.456	-.464	-.217
Weather=Raining	.210	-.367	.550	.015	-.113	-.088	.357	-.566	-.063	.053	.262	.235	.351	.270
Weather=Snowing	-.489	.489	.344	.399	-.484	.204	.122	.447	-.300	.377	-.089	.359	-.336	.383
Involved vehicles=1	.015	-.522	.443	.224	-.343	-.015	.289	-.192	-.480	-.140	.139	-.213	.431	-.228
Involved vehicles=2	.216	-.451	-.041	.261	-.046	.107	.245	.184	-.394	.131	-.459	.203	-.256	.434
Involved vehicles=3	-.394	.397	-.332	.345	.116	-.110	.341	.182	-.093	.156	.477	-.258	-.011	-.492
Vehicle Type=Bicycle	.421	-.282	.172	-.509	.199	.168	-.222	-.240	-.412	.034	-.286	.265	.169	.488
Vehicle Type=Car	-.448	-.499	.100	-.114	.019	.088	-.225	.600	.011	-.587	-.246	-.101	-.374	.412
Vehicle Type=PTW	.207	.043	-.116	.111	.109	-.059	.136	-.123	-.204	-.176	.415	-.429	-.110	-.174
Vehicle Type=Truck	.319	-.611	-.726	-.221	-.006	-.312	-.205	.597	.168	-.225	.269	.289	.483	-.405
Vehicle Age=>20	.092	-.095	.234	-.021	-.199	.023	-.477	.425	-.208	.371	.192	-.315	-.461	.456
Vehicle Age=0-10	-.305	.110	.441	-.037	-.276	.352	.005	-.098	.061	.069	-.308	-.243	-.192	.116
Vehicle Age=10-20	.502	-.335	.312	-.222	-.253	.393	-.066	-.001	-.518	-.399	-.090	-.478	-.396	-.145
Driver Behaviour=Disobeying pedestrian crossing facility	.123	.042	.428	.250	.268	.129	-.389	.300	.255	.141	.036	-.032	-.462	-.343
Driver Behaviour=Disobeying stop sign	-.413	.358	-.076	-.006	.328	-.429	.269	.145	-.058	.206	.274	-.259	-.095	.206
Driver Behaviour=Distraction	.038	-.402	-.199	-.353	.214	.220	-.445	.192	-.478	-.420	-.222	.421	-.375	.358
Driver Behaviour=Illegal travel direction	-.390	.128	.414	-.350	-.078	.466	-.089	.137	.027	-.199	-.154	.259	-.223	-.120
Driver Behaviour=Manoeuvering	-.296	.452	.294	-.173	-.366	.023	-.011	-.295	.147	-.212	.283	-.419	-.340	-.272
Driver Behaviour=Normal	-.317	-.006	-.352	-.285	.126	-.381	-.498	-.325	.035	.330	.048	-.373	.035	.141
Driver Behaviour=Speeding	.488	-.135	.134	-.007	-.422	-.394	.254	.415	-.100	.491	-.120	-.401	.298	-.439
Driver Behaviour=Tailgating	-.016	.265	.297	.068	-.349	-.343	-.100	.355	-.380	-.293	.200	-.367	.396	-.345
Vehicle Defect=Defect	-.428	.228	-.186	.049	.000	-.267	.417	-.375	-.406	.164	.377	.015	.081	.375
Vehicle Defect=No defect	-.288	.236	-.221	.488	-.048	-.405	-.337	.101	.040	-.184	-.028	.176	-.510	.074
Driver Psychophysical State=Alcohol	.114	-.333	-.446	.232	.398	-.002	.486	.198	-.146	.308	-.421	.040	-.065	.120
Driver Psychophysical State=Dazzled	-.178	-.126	.217	-.217	-.386	.331	.376	-.417	.267	.424	-.391	.463	-.420	.265
Driver Psychophysical State=Drug	-.144	-.530	.076	-.361	-.162	.046	-.244	.008	-.003	-.289	-.066	.213	.352	.334

Predictor	Predicted													
	Hidden Layer 1													
	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)	H(1:14)
Driver Psychophysical State=Exceeding the prescribed driving period	.243	.288	-.026	-.340	-.338	-.466	-.042	.442	-.225	.080	-.085	.217	.362	-.049
Driver Psychophysical State=Illness	.277	.288	.318	.097	.113	-.374	.262	-.322	.083	-.232	.468	.070	.299	.124
Driver Psychophysical State=Normal	.313	-.105	.443	.441	.282	-.060	-.125	-.282	-.426	.374	-.021	-.278	-.404	.093
Driver Psychophysical State=Sleeping	-.419	-.302	.420	.338	.242	.125	-.459	.313	.169	-.085	-.008	-.168	-.271	.473
Driver Psychophysical State=Uncorrected, defective eyesight	.404	-.234	.448	.460	.170	.234	.018	-.421	.008	-.247	.096	-.272	.093	.068
Driver Age=0-17	.185	.087	.187	-.063	.178	.008	.123	.364	.069	.094	-.327	.069	.186	.107
Driver Age=18-24	.513	-.306	.089	.354	-.395	-.459	.012	.272	-.318	.608	.210	-.031	-.303	-.023
Driver Age=25-44	-.491	.229	-.404	-.373	.161	-.477	.002	.402	.048	.320	-.184	.217	-.018	.222
Driver Age=45-54	.084	-.132	.281	.038	-.228	.426	.334	-.097	-.383	.426	.082	.445	-.181	-.214
Driver Age=55-64	.354	.079	.458	.342	.232	.002	.521	.494	.064	-.069	-.004	-.249	.106	.196
Driver Age=65-74	.186	-.402	.246	.163	-.049	.045	-.095	.234	.169	-.273	.084	.056	.155	-.488
Driver Age≤75	.102	-.384	-.118	-.172	.395	.167	.159	.128	.450	.162	.152	-.489	-.267	-.334
Driver Gender=Female	.492	.170	-.120	.343	.171	-.372	-.235	-.348	.158	-.520	.294	-.073	-.154	.353
Driver Gender=Male	-.089	.151	-.172	.261	.325	.291	-.321	.396	.312	.120	-.570	-.191	.438	-.416
Pedestrian Behaviour=Crossing on pedestrian crossing facility	.050	-.133	-.051	-.336	.203	-.084	.250	-.177	.004	.273	-.170	-.422	-.363	.039
Pedestrian Behaviour=No Crossingon pedestrian crossing facility	.114	-.043	.157	-.090	.154	.442	.524	.338	-.314	-.037	.287	.481	.291	.112
Pedestrian Behaviour=Walkingfacing the traffic	.477	.386	.275	.338	-.200	.276	-.442	.049	-.451	.178	-.185	.358	.043	.161
Pedestrian Behaviour=Walking back to the traffic	-.214	.239	.129	-.071	-.025	-.336	.414	.033	-.152	.061	-.073	-.460	-.145	.193
Pedestrian Behaviour=Walking Regularly	.453	.130	-.354	.336	-.253	.342	-.086	-.101	.327	-.113	.388	-.091	.093	-.049
Pedestrian Psychophysical State=Alcohol	-.232	-.271	-.132	-.447	-.176	.129	-.369	-.257	-.377	.298	-.076	-.431	.028	.124
Pedestrian Psychophysical State=Drug	-.266	.430	.144	.161	.223	-.173	-.258	.393	.184	.112	.276	-.185	-.347	-.139
Pedestrian Psychophysical State=Illness	-.053	.263	.326	.044	.058	-.083	.149	.344	-.476	-.368	.321	.381	-.382	-.189
Pedestrian Psychophysical State=Normal	-.401	-.018	-.165	.403	.441	.461	.315	-.307	-.219	.211	.304	.496	-.015	.411
Pedestrian Gender=Female	-.473	-.188	-.070	.348	.277	.380	-.293	.121	.130	.113	.251	-.212	-.563	-.005
Pedestrian Gender=Male	-.410	.205	-.596	.311	.340	-.417	-.131	.501	.248	-.043	-.342	-.077	-.032	.257
Pedestrian Age=0-14	-.278	-.202	-.169	.142	.250	.361	-.053	-.735	.500	-.011	-.471	-.143	-.340	.065

Predictor	Predicted													
	Hidden Layer 1													
	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)	H(1:13)	H(1:14)
Pedestrian Age=15-24	.282	.519	.061	.356	.562	-.037	-.511	-.168	-.096	-.080	.101	-.459	.075	.019
Pedestrian Age=25-44	-.256	.411	.157	.014	-.076	.305	.066	-.406	.456	-.272	-.392	-.211	.221	-.361
Pedestrian Age=45-54	.203	.014	-.297	-.319	-.301	.331	-.361	-.306	-.198	.001	-.231	-.023	-.381	.074
Pedestrian Age=55-64	-.398	.155	-.500	.426	.227	-.521	.059	-.271	-.345	.469	.128	.146	.270	-.464
Pedestrian Age=65-74	-.258	-.034	-.295	.123	.202	-.464	-.159	.423	.365	.745	-.325	-.293	.768	.263
Pedestrian Age≤75	-.289	-.880	-.344	.647	-.322	-.374	.628	1.221	-.206	.694	.105	-.201	-.261	.030

Table 177 – Artificial Neural Network parameter estimates for the output layer, Italy.

		Predicted	
		Output Layer	
Predictor		**Injury**	**Fatal**
Hidden Layer 1	(Bias)	-0.160	-0.146
	H(1:1)	-0.561	-0.415
	H(1:2)	0.460	-0.304
	H(1:3)	-0.004	-0.306
	H(1:4)	0.003	0.458
	H(1:5)	0.484	-0.168
	H(1:6)	0.501	0.056
	H(1:7)	-0.253	0.280
	H(1:8)	-0.523	0.562
	H(1:9)	-0.032	0.108
	H(1:10)	-0.633	0.197
	H(1:11)	0.105	-0.007
	H(1:12)	-0.298	0.037
	H(1:13)	-0.308	0.390
	H(1:14)	-0.093	-0.039