Risk analysis of bicycle accidents: A Bayesian approach

Zaili Yang[1], Zhisen Yang[1*], John Smith[2], and Bostock Adam Peter Robert[3]

1. Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool, UK
2. Merseytravel, Liverpool, UK
3. Mersey Police, Liverpool, UK

*Corresponding author: Zhisen Yang

Email address: zhiseny@163.com

Liverpool Logistics, Offshore and Marine (LOOM) Research Institute

Department of Maritime and Mechanical Engineering

Liverpool John Moores University

Highlights

- Hazards influencing cycling safety are identified from previous literatures and accident reports.
- A new conceptual risk analysis and prediction model based on a Bayesian network is developed to enable the analysis and prediction of cycling accident severity.
- The influence of single hazard and the combination of multiple hazards on accident severity are evaluated.
- Safety suggestions and an early-warning system are provided to both transport authorities and cyclists to help them reduce the severity of possible cycling accidents and ensure cycling safety.

Risk analysis of bicycle accidents: A Bayesian approach

Abstract

Cycling helps reduce traffic congestion and environmental pollution and promote a healthy lifestyle for the general public. However, it could also expose cyclists to dangerous environments, resulting in severe consequences and even death. Transport authorities are seeing growing accidents in city regions with increasing cycling population, requiring the development of new risk informed cycling safety policies. This paper aims to develop a new conceptual risk analysis approach based on a Bayesian network (BN) technique to enable the analysis and prediction of the severity of cycling accidents. To identify the risk factors influencing cycling accident severity, 2,000 cycling accident reports from the UK city region were manually collected, where primary data was extracted and analysed and an advanced data training method (i.e. Tree Augmented Naïve Bayes (TAN)) was applied to find their correlation and use a BN to investigate their individual and combined contributions to cycling accident severity. As a result, the risk factors influencing accident severity are prioritised in terms of their risk contribution. The risk levels of accident severity can be predicted and analysed in dynamic situations based on the data from simulated and/or real cycling environments. The findings can provide useful insights for making rational cycling safety policies in proportion to different risk levels.

Keywords: Cycling safety, Bayesian network, accident severity, transport risk analysis

1. Introduction

Cycling (which was once a neglected and unvalued transportation mode) is becoming a popular way for mobility, recreation, exercise and sports worldwide. The number of bicycles in use reached an estimation of 800 million in 2004, twice the number of cars (Peden et al., 2004) and 580 million bicycles were in private household ownership (Oke et al., 2015). According to the Walking and Cycling statistics published by Department for Transport of United Kingdom (UK), the average number of miles cycled per person in 2019 (54 miles) has generally increased over 40% since 2002 (39 miles), and such trend is growing at a steady pace. Meanwhile, the cost spent on bicycles and bicycle equipment is around £35 million in UK in 2019, while it was only £24 million in 2013. Additionally, according to the world cycling index released by Eco Counter in 2019, the global bicycle traffic is experiencing a consecutive increase in recent years, 8% from 2013-14, 3% from 2014-15, 0.5% from 2015-16, 0.2% from 2016-17, and 6% from 2017-18. In fact, it is not surprising to observe an increase in popularity of cycling as it has been treated as a way to reduce traffic congestion and environmental pollution, and promote a healthy lifestyle for the public (Anderson et al., 2000; Higgins, 2005; Heinen et al. 2010; Heydari et al., 2017). Other benefits cycling brings to individual users, includes the easiness of the burden of vehicle parking, exercise of the body and reduction of travel cost. As a result, the growing popularity in transport cycling triggers the increasing interest of city councils in making urban transport infrastructure more bicycle friendly. For instance the pioneering cities such as Amsterdam and Copenhagen, are enjoying the benefits of their efforts, with 40% of trips being completed by bicycles (Pucher et al., 2010).

Despite such benefits, the safety of cycling is under debate due to the vulnerable nature of cyclists and a broader age distribution from children to the elderly compared to the other types of road users. In many countries of mixed traffic systems, cyclists often have to use the same infrastructure as cars, buses and trucks but are not protected like motorised road users (Reynolds et al., 2009). These adverse natures and conditions could expose cyclists in dangerous environments. Based on the official mode-by-mode fatality and travel statistics of the US Department of Transportation (National

Highway Traffic Safety Administration and Federal Highway Administration), bicyclists were 12 times more likely than car occupants to be killed (72 vs 6 fatalities per billion kilometres (Pucher and Dijkstra, 2003). Once cycling accidents happen, they could result in severe consequences, including major injuries, deaths and economic loss due to the traffic blocks they cause. Therefore, how to reduce the risk and consequence of accidents that cyclists may encounter and improve cycling safety becomes an urgent research problem to be addressed. They are forming a major concern for many transportation authorities in large cities in the world, leading to a substantial growth in cycling safety research. Although showing increasing concerns, scientific risk analysis and safety management using advanced uncertainty modelling technologies on cycling safety is still scanty in the literature, particularly compared to other transport modes.

To fulfil this research gap, this paper aims to develop a new conceptual risk analysis approach based on a Bayesian network (BN) to enable the analysis and prediction of cycling accident severity using the data derived from 6-year (2012-2018) transport accident reports involving cycling in the Liverpool city region. To develop the BN-based risk model, a review on the related works on cycling safety is conducted to identify the risk factors influencing the severity of cycling accidents worldwide. The risk factors are classified into different categories, e.g. cyclist behaviour and personal factors, environmental conditions, road facility issue, interaction with other road users, hazardous road conditions and bike-related factors. The factors that are most frequently discussed and analysed are selected for further investigation in this research work.

Next, all the road accident reports involving cycling in the Liverpool city region from 2012-2017 were preliminarily analysed to derive the initial primary risk data. Such primary accident data are incorporated to verify the identified major cycling risk factors. Through statistical analysis, the interdependence of the risk factors and their joint effect on accident severity are obtained and used as the input to a risk analysis and prediction model for supporting cycling safety policy making. The BN model is able to analyse the key risk factors influencing cycling accident severity. In addition, empirical cases based on the new set of data collected from the accidents reported in 2018 are used to validate the BN model and its prediction accuracy in various hazardous situations and generate useful insights for accident prevention. Based on the findings, safety suggestions are provided to both transport authorities and cyclists to help them reduce the severity of possible cycling accidents.

The novelty of this research lies in the following aspects: 1) The risk factors influencing cycling accident severity are identified from the combination of the related literatures and real historical accident statistics. 2) The big data are collected and processed for the development of a data-driven BN risk model, containing more 200,000 pieces of risk information with regards to over 100 different risk parameters. 3) It enriches the quantitative cycling risk analysis literature by incorporating advanced uncertainty modelling (e.g. BN). 4) The accuracy and robustness of the risk prediction model are tested using a new set of data collected from 2018. The model, capable of accurately predicting the risk severity in over 95% real cases, can provide useful insights for policy making.

The remainder of this paper is organised as follows. Section 2 reviews the current literature relating to cycling accidents to identify related hazards and risks, as well as a discussion on the papers relating to cycling safety using risk assessment approaches. Section 3 describes the methodologies and techniques applied in this study, which is followed by the risk-based cycling model construction and verification in Section 4. In Section 5, the sensitivity analysis is conducted through a two-step approach for drawing useful findings in terms of the severity of cycling accidents based on a real case of the Liverpool city region. Finally, Section 6 concludes this study with reference to its contributions and implications.

2. Literature Review

2.1 Hazards in cycling safety

Previous studies on cycling safety focused on a wide range of hazards influencing the occurrence probability and/or consequence severity of accidents. A combination of keywords ‘cyclist’ and ‘risk analysis’ is used when searching in Web of Science, resulting in 250 related papers. Based on the following criteria, a relevant literature database was established. The first criterion is that any paper focusing on the analysis of crash and injury rates, medical care are excluded. Secondly, book chapters, papers written in other languages, and papers lacking basic information are excluded.

As a result of these filters, 100 relevant papers from year 1990 to 2017 are systematically reviewed. These hazards are classified into six categories based on their features, including 1) Cyclist behaviour and personal characteristics; 2) Environmental conditions; 3) Road infrastructure issue; 4) Interaction with other road users; 5) Hazardous road conditions; 6) Bike-related factors

The identified hazards and their appearance frequencies of each category in the literature can be found in Appendix 1.

2.2 Risk assessment studies in cycling safety

Since cycling safety has received increased attention from the public, researchers and transport authorities, there is a growing profile on the relevant studies in the literature (i.e. Osama & Sayed, 2017). However, most of the previous studies are conducted to analyse and evaluate risks brought by a selected hazard leading to cycling accidents. In other words, little research investigates the risk assessment study of cycling involving multiple hazards, from a practical point of view. They focus more on risk analysis and diagnosis using traditional risk analysis methods than risk prediction using multiple risk influencing factors, from a methodological perspective.

Using the data from Los Angeles, Behnood & Mannering (2017) applied a random parameters multinomial logit model to estimate the effects of a wide range of variables on accident consequence severity. A comprehensive analysis of the influence of multiple risk variables on accident severity was presented. Based on the ultrawideband (UWB) technology, Dardari et al. (2017) proposed an UBWlocalization system to improve the safety of cyclists. Combined with enhanced risk assessment units, the peculiarities of the system in terms of accuracy and cost enable a real-time warning function to road users. The system is proved efficient and safe to use for cyclists. Realising the potential of mental mapping in recording and analysing safety perceptions in cycling safety, Manton et al. (2016) developed a novel method to model the individual and structural determinants of perceived cycling risk through the derived qualitative and quantitative data. Through investigating a real case in Galway City, it was found that the proposed model is useful for assessing the perceptions of cycling risk with a strong visual aspect and improving public transportation safety significantly. Focusing on the Brussels-Capital region, Vandenbulcke et al. (2014) utilized a spatial Bayesian modelling approach to predict cycling accident risk of a whole transportation network, as well as identify how cycling safety is influenced by road infrastructure. Considering infrastructure, traffic and environmental characteristics, the research reveals several occasions of high cycling risks, such as bridges without cycling facility and complex intersections. Such findings are helpful for local transport authorities to predict the accident risk when making policies and for cyclists to choose the safest routes. It explores a new research direction on safe cycling by employing advanced uncertainty modelling like BNs, requiring the integration of more experimental work with rich data and more influencing factors before its wide applications in practice.

Taking advance of casual inference, BN can be used to analyse the importance degree of risk factors and the relationships among them. Compared to pure Bayesian theory, BN is more visualized. Furthermore, compared to other graphic models, it has a foundation of mathematical knowledge. BN is also widely used to evaluate and predict risks in various transportation mode because of its advantages in forward prediction analysis and backward risk diagnosis (e.g. Zhang et al., 2013), Xie et al., 2007; Krause et al., 2016; Yang et al., 2017; Hänninen & Kujala, 2012; Li et al., 2014; Yang et al., 2018a; Serrano et al., 2018; Yang et al., 2018b). However, BN's applications in cycling safety is scanty (e.g. Puchades al., 2018; Kondo et al., 2018; Chen, 2016) and its ability of forward prediction analysis and backward risk diagnosis is yet to be sufficiently explored in cycling safety. This is largely due to the lack of real accident data and complicated dependence among the involved risk factors. This study will pioneer the use of data-driven BNs to analyse the severity of cycling accidents for transport safety policy making.

2.3 Data-driven approaches for BN structure learning

The structure of BN risk models is often developed through human knowledge. Despite this, a common criticism is that such an approach is time consuming and heavy emphasis is placed on experts to provide both the local probability parameters and dependence among the parameters, which often introduces subjective bias into the model. An alternative method for BN construction is to induce the network structure from data, namely the data-driven approach.

The search and score approach, (which is widely applied in this field) seeks to explore a search space of candidate BN structures for the one that best represents the causality and dependency relationships (Cooper et al. 1992). Cooper and Herskovits (1992) derived a K2 scoring metric based on Bayes theorem, starting with an empty network and iterating through each node to get the best structure. An order among the variables needs to be assumed in this algorithm, which makes it hard to determine. In contrast to Cooper and Herskovits, Buntine's 'B' algorithm (1991) does not require a variable order. A link will be added at the end of each iteration if it can maximize the score and does not lead to a cycle, until the score no longer increases. However, once local optima occur, the algorithm could not give reasonable results.

In recent years, Naïve Bayes (NBN) learning has been developed as a popular network construction algorithm. It can reduce the construction complexity given that the parameter learning in the model do not need complicated iteration process, as well as effectively avoid the subjectivity of expert judgment (Wang et al., 2018). However, the assumptions of building NBN is sometimes too strong to be realistic. In order to improve the performance of NBN, its structure is augmented with links among the attributes or factors. This type of structure that does not require independence among attributes is called augmented BN (ABN). Further, if the class variable has no parents, and each attribute has the class variable and at most one other attribute as parents, the ABN under this condition is called Tree augmented Naïve Bayes (TAN). Compared to other data-driven network construction approaches, like naive BN (Langley et al., 1992) and C4.5 (Quinlan, 1993), TAN is proven to be more competitive and accurate (Murphy et al. 1995). Due to such strengthens, its wide applications are spread across different risk studies in the transportation field (e.g. Yang et al., 2018; Wang and Yang, 2018).

3. Methodology for model construction

To develop the data-driven BN model for the analysis and prediction of cycling accident severity, a conceptual methodology consisting of four steps is developed in this section, including data acquisition, variable identification, structure learning, and model validation. Before the model construction, a basic assumption lies that all the variables in the model are conditionally independent given the value of the target node.

3.1 Data acquisition

The data used to construct the BN comes from the STATS19 accident reports applied by the Department of Transport in the UK. It consists of a set of data that are collected by a police officer when a road accident is reported. The accident reports are used extensively for research work and for guidance in the improvement of road safety policies in relation to road, road users and vehicles.

In this paper, 2258 STATS19 reports involving cycling accidents in Liverpool from 2012-2017 are collected from Merseyside Police as the case data to support the development of the proposed BN in the first round of model development. Each report includes a great variety of characteristics related to the cycling accident (e.g. day, time, road surface, weather, lighting) involving more than 100 parameters influencing the accident severity, as seen in the official STATS19 form from UK DfT website (UK DfT STATS19 form, 2020).

3.2 Variable identification

Given the fact that many factors are irrelevant to any of the collected reports, a screening process is conducted to only retain the relevant parameters for the development of the BN based cycling risk model and to gather information on the definition of the grades of the employed variables. A simple flowchart is presented as follows to display this process – see Figure 1.

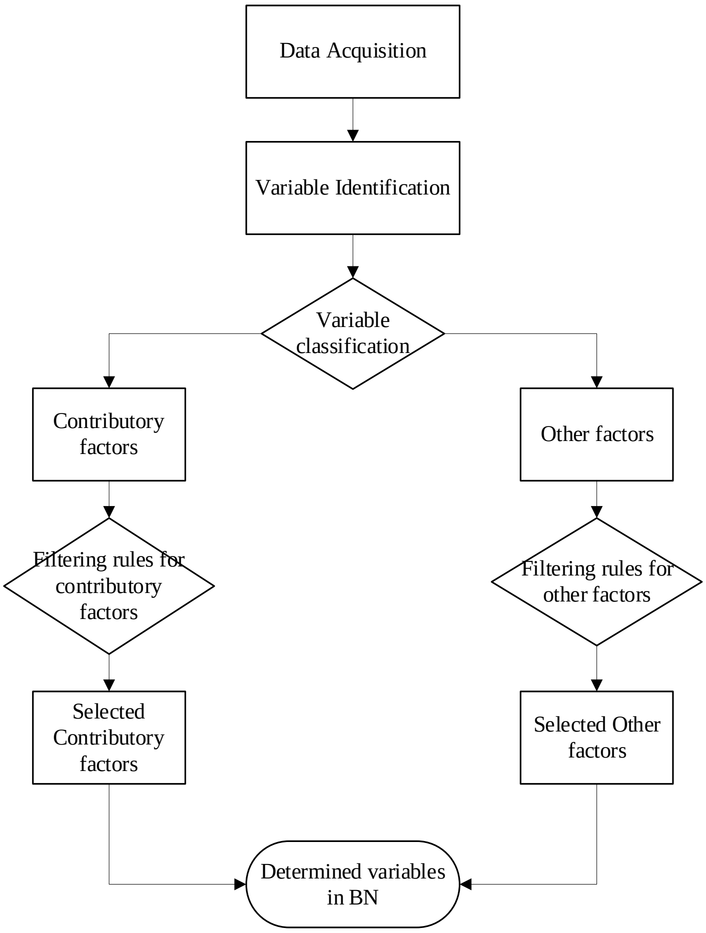

Data Acquisition
Variable Identification
Variable classification
Contributory factors
Other factors
Filtering rules for contributory factors
Filtering rules for other factors
Selected Contributory factors
Selected Other factors
Determined variables in BN

Figure 1. Flowchart of variable determination process

The description of the proposed variable identification and screening process is outlined in the following steps:

Step 1. Data collection and acquisition

Step 2. Variable identification from collected accident reports

Step 3. Variable classification based on the identified variables in step 2 in two groups: contributory factors and other factors

Contributory factors

In a STATS19 report, there is an important component called ‘contributory factors’, which reflects the reporting officer’s opinion at the time of reporting. The contributory factors are largely subjective and depend on the skill and experience of the investigation officer to reconstruct the events that directly lead to the accident. The identified factors for the accidents are clarified based on the evidence rather than subjective judgments of the officers on duty about what may have happened. It is an alternative form of expert judgment on the key actions taken by the involved driver(s) and cyclist(s) and the failures that directly lead to the accidents, presenting a valuable variable in predicting accident severity. These factors are defined from *STATS20 handbook* issued by Department for Transport of UK (UK DfT STATS20, 2020).

In total, there are 78 contributory factors. Against the same factor, an accident is reported from two different types of involving users as suggested by the STATS20 handbook: victim and the encountering road users. Although the number of involving people in an accident varies, they all belong to these two types. Therefore, in our model construction, the selected contributory factors consist of two aspects: contributory factors of victim and contributory factors of other encountering road users.

Step 4 Variable screening through corresponding filtering rules.

Step 5 Determination of final selected variables used for model construction.

The detailed information of the screening process is found in Appendix 2.

3.3 BN structure learning through a TAN approach

A BN structure is learnt by two means including through subjective expert knowledge and objective data training (i.e. data driven BN). The subjective approach is normally time consuming, and heavily relies on the domain experts to provide both the local parameters’ probabilities and global dependency among the parameters, causing the concerns on the model’s robustness and result’s accuracy. An alternative method for the BN construction is to induce the network structure from objective data, namely the data-driven approach, which can greatly reduce the subjective bias and increase the soundness of the model (Oteniya, 2008).

Different data training approaches have been used to learn BN structures. In this study, TAN learning is adopted to learn the structure of the BN for cycling safety analysis. TAN learning is a semi-naïve Bayesian learning method. It relaxes the naive Bayes attribute independence assumption by employing a tree structure, in which each attribute only depends on the class and one other attribute. Compared to other approaches (e.g. naïve BN, C4.5), most of which have a local optimal problem when generating a BN structure, the TAN learning is more efficient and accurate. Based on the experiments carried out by University of California at Irvine (UCI), TAN learning has revealed significant improvement over other approaches in terms of model accuracy. TAN learning not only maintains the robustness and computational complexity of Naïve BN learning, but also displays better result accuracy (Friedman et al., 1997). Based on the risk variables identified in Appendix 1, the quantitative BN to represent the interactive dependencies can be constructed through the TAN learning as follows.

3.3.1 TAN learning

The essence of TAN learning is actually an optimization problem. Let $A_1, \dots, A_n$ be the attribute variables (the influencing variables in Section 3.2 like ‘District’, ‘Time’, ‘Weather’, etc.) and C be the class variable (target variable 'Accident severity') in the analysis of cycling accident severity. Π_C represents the parent variables of C. B is defined as a TAN model if $\Pi_C = \emptyset$ and there is a function π that defines a tree over $A_1, \dots, A_n$ such that $\Pi_{A_i} = \{C, A_{\pi(i)}\}$ if $\pi(i) > 0$, and $\Pi_{A_i} = \{C\}$ if $\pi(i) = 0$. The optimization problem consists of finding a tree defining function π over $A_1, \dots, A_n$ such that the log likelihood is maximized, and the TAN model under this function is used as the structure of the target BN model. One difference between a traditional BN model and the TAN model lies in class variables. Class variables in the BN model always have at least one parent node. However, since Bayesian inference will be used on the results, it is accepted for links to go in either direction to fit the result reflecting the reality. In other words, the directions of links in the TAN model can be changed appropriately to fit the demand of this study on cycling safety.

The procedure entitled 'Construct-TAN' can solve the above optimization problem. This procedure follows the general outline proposed by Chow and Liu (1968), except that instead of using the mutual information between two attributes (i.e. any two of the factors in Appendix 2), it uses conditional mutual information between attributes given the class variable (i.e. accident severity). This function is defined in Equation 1.

$$I_P\left(\boldsymbol{A_i}; \boldsymbol{A_j} \middle| \boldsymbol{C}\right) = \sum_{\boldsymbol{a_{ii}}, \boldsymbol{a_{ji}}, \boldsymbol{c_i}} P\left(\boldsymbol{a_{ii}}, \boldsymbol{a_{ji}}, \boldsymbol{c_i}\right) log \frac{P(\boldsymbol{a_{ii}}, \boldsymbol{a_{ji}} | \boldsymbol{c_i})}{P(\boldsymbol{a_{ii}} | \boldsymbol{c_i}) P(\boldsymbol{a_{ji}} | \boldsymbol{c_i})} \quad (1)$$

where I_P represents the conditional mutual information, $\boldsymbol{a_{ii}}$ is the *i*th state of the attribute variable $\boldsymbol{A_i}$, $\boldsymbol{a_{ji}}$ is the *i*th state of the attribute variable $\boldsymbol{A_j}$, $\boldsymbol{c_i}$ is the *i*th state of the class variable $\boldsymbol{C_i}$. This function measures the information that both $\boldsymbol{A_i}$ and $\boldsymbol{A_j}$ have when the value of $\boldsymbol{C}$ becomes known.

The Construct-TAN procedure for cycling accident severity analysis and prediction consists of five main steps:

a) Compute $I_P\left(A_i, A_j \middle| C\right)$ between each pair of attribute variables in cycling safety, $i \neq j$.
   Attribute variables in this research: all the influencing variables identified in Appendix 1.
   Class variables in this research: Accident severity.
b) Build a complete undirected graph in which the vertices are the attributes $A_1, \dots, A_n$.
   Annotate the weight of an edge connecting A_i to A_j by $I_P\left(A_i, A_j \middle| C\right)$.
c) Build a maximum weighted spanning tree.
   A spanning tree is a connected subgraph containing no cycles. The maximum weight-spanning tree is a spanning tree, compared to which no other spanning tree has a larger sum of weights on its edges. Therefore, the maximum weighted spanning tree in this paper is the tree that has a maximum sum of $I_P\left(A_i, A_j \middle| C\right)$.
d) Transform the resulting undirected tree to a directed one by choosing a root variable from the attribute variables and setting the direction of all edges to be outward from it.
e) Construct a TAN model by adding a vertex labelled by the class variable *C* and adding an arc from C to each A_i.
f) Estimate the conditional probability of each variable/node through a gradient descent approach (Yang et al., 2018)

3.4 Model verification

To verify the proposed model, new accidents happened in the Liverpool region in 2018 are collected from Merseyside Police. If the estimated results delivered by the proposed BN model are keeping a high harmony with the real results of the new accidents, the model is proven robust; otherwise, the model fails.

3.5 Sensitivity Analysis

Sensitivity analysis is known as a way to determine how the uncertainty in the output of a model can be influenced by the different sources of uncertainty in its input. In this particular study, a two-step sensitivity analysis has been developed to determine the influence degree of risk variables in Appendix 1 on accident severity. The findings will provide useful insights for transport authorities for developing their cycling safety policies.

3.5.1 Mutual information

Mutual information (entropy reduction) is a quantity that measures how much one random variable tells about another. High mutual information indicates close connection; low mutual information indicates weak connection; zero mutual information indicates two variables are independent. To understand the nature of mutual information, entropy needs to be initially defined. The higher the entropy, the more uncertainty one is about a random variable. Shannon (1949) indicated that the measure of uncertainty of a random variable should be a continuous function of its probability distribution and should satisfy some specific conditions:

1) It should be maximal when its probability distribution is uniform, and in this case, it should increase with the number of possible values the variable can take;
2) It should remain the same if we reorder the probabilities assigned to different values of the variable;
3) The uncertainty about two independent random variables should be the sum of the uncertainties about each of them.

Based on these conditions, the entropy is defined. Consider a discrete random variable $\boldsymbol{\alpha}$ with possible values $\{\alpha_1, \alpha_2, \dots, \alpha_i\}$ and probability distribution function $P_\alpha(\boldsymbol{\alpha})$, then the entropy can be explicitly written as (Yang et al., 2018):

$$H(\boldsymbol{\alpha}) = -\sum_i P_\alpha(\alpha_i) log_b P_\alpha(\alpha_i) = -E_P log_b P(\alpha)$$

where b is the base of the logarithm used. Normally, the value of b is 2; E_P is the expected value over the probability distribution.

Further, the conditional entropy is the average uncertainty about one variable $\boldsymbol{\alpha'}$ after observing a second random variable $\boldsymbol{\alpha}$, and is given by:

$$H(\boldsymbol{\alpha'}|\boldsymbol{\alpha}) = \sum_i P_\alpha(\alpha_i)\left[-\sum_{i'} P_{\alpha'|\alpha}(\alpha'_i|\alpha_i) log_b P_{\alpha'|\alpha}(\alpha'_i|\alpha_i)\right] = E_{P_\alpha}\left[-E_{P_{\alpha'|\alpha}} log_b P_{\alpha'|\alpha}\right]$$

Based on the definition of entropy, conditional entropy and mutual information, it is easily to find out that mutual information is the reduction in uncertainty about a variable. Assuming S represents 'accident severity', β represents a random risk variable, β_i represents the ith state of β, $I(S,\beta)$ represents the mutual information between 'accident severity' and risk variables. $I(S,\beta)$ can be written as in Equation 2

$$I(S,\beta) = H(S) - H(S|\beta) = -\sum_{d,i} P(s,\beta_i) log_b \frac{P(s,\beta_i)}{P(s)P(\beta_i)} \qquad (2)$$

In this paper, the introduction of mutual information is to measure the mutual dependence of different risk variables influencing safe cycling. In other words, it is the information that two variables share. It is the value used to measure the strengths of the relationships between the target node (i.e. accident severity) and influencing nodes (i.e. day, time, weather, road surface). The larger the value of mutual information, the stronger relationship that exists between the risk variable 'β' and 'accident severity'. The factors having stronger relationships with 'accident severity' are viewed as significant variables in cycling safety. One of the advantages of mutual information is that it can be computed between the variables at different layers. When a new observation of an influencing variable is obtained, the mutual information can help measure the uncertainty of the observation on target node (i.e. accident severity).

3.5.2 Combined influence of multiple variables

The value of mutual information can measure the significance and influence degree of individual variables. If the mutual information is low, this variable does not have strong relationship with the target node ‘accident severity’. However, sometimes these variables of insignificant impact on ‘accident severity’ will generate a much higher effect in a combined way. Therefore, a further sensitivity analysis focusing on the combined influence of the investigated risk variables are conducted to find the combined sets of influencing variables that can generate significant impact on accident severity.

4. Model construction, results & verification

4.1 Description of nodes in the BN

In this section, the risk variables influencing the accident severity after screening process are explained with a particular reference to their state definitions.

4.1.1 Influencing variables

1) District

This refers to the region/location where cycling accidents happen. According to the STATS19 reports, there are five major districts in the investigated Liverpool city region: Knowsley, Sefton, City of Liverpool, St. Helens and Wirral.

2) Day

The statistics of cycling accidents reveals that the frequency of cycling accidents and their severity varies from days to day. For example, Sunday has the lowest number of accidents in total, while Tuesday has the most severe or fatal accidents. This node is therefore discretised into seven states: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday.

3) Time

This variable is classified into two states based on: rush hour or not. According to BBC News and Wikivoyage, rush hour in the UK is typically 7am-10am and 4pm-7pm every weekday; for weekends, there is no specific definition. In this study, '11am-7pm' is set as the rush hour on weekends in Liverpool region because the traffic volume during this period is significantly higher than other time in weekends. Two states are set as rush hour and non-rush hour.

4) Encountering vehicle types

Encountering vehicle is the other party colliding with a cyclist on the road. Different encountering vehicle types often result in different accident consequences. For example, it is undoubtedly that colliding with a heavy goods vehicle (HGV) is much more dangerous than the collision with a motorbike for a cyclist if other conditions are kept the same (i.e., speed, environment). Based on the information provided by the accident reports, this variable has six states: Cars, HGV, Public Service Vehicle (PSV), motorcycle, cyclist, and other/unknown.

5) Weather

This variable refers to weather conditions at the time and location of a cycling accident. As stated in Section 2, bad weather conditions have a major impact on cycling safety and failure to recognize its impact may cause huge loss, injuries and even casualties. The effect of a bad weather condition on cycling safety is mainly because of the reduction in visibility and distraction of cyclists, as well as its impact (e.g. rain) on a hazardous road condition (Joon-Ki Kim et al., 2007). Therefore, this variable needs to be paid much attention, especially in regions where bad weather often occurs. The Department for Transport in the UK defines several states for this variable: fine with high winds, fine without high winds, rain with high winds, rain without high winds, and others.

6) Road surface condition

The road surface condition in this study refers to the surface condition at the time and the place of the cycling accident. According to STATS19 reports, there are five types of a road surface condition in the UK: dry, wet/damp, snow, frost/ice, and flood (where surface water is over 3cm deep). However, in the Liverpool region, the major road surface conditions are dry and wet/damp, as stated by Merseyside Police. Meanwhile, the cycling accident database also tells there are very few accidents occurring on snow/frost/ice/flood road surface and none of them causes fatal consequence. Hence, this variable is classified into three states: dry, wet/damp, and others (i.e. snow, frost, ice and flood).

7) Street lighting

Darkness is the most mentioned environmental hazard for cyclists, according to the literature. Previous researchers have already shown that cycling during late hours, especially at night, is more hazardous than daytime (Juhra et al., 2012). As a solution, the use of sufficient street lighting facilities can effectively tackle the darkness issue. However, in Liverpool, not all the places have the street lighting facilities, or some facilities are broken and not working well, generating an impact on the accident severity accordingly. Based on the STATS19 reports, ‘street lighting’ is categorized into three states in this study with respect to the darkness types: dark with no street light/unknown, dark with street lights present and lit, and daytime.

8) Combined road class

When a road accident happens at a junction, the police officer will record the main road class and the second road class in the STATS19 report. The main/first road is defined as the one with the highest class of all the roads entering the junction, and the second-class road is the second highest one. Different combinations of road classes represent different environments, which is an important factor affecting the accident severity, named as ‘combined road class’. It is denoted as the ‘1st road class - 2nd road class’.

Normally, there are five road classes in the UK: A, B, C, M, and other unclassified. In this research, C and M road classes are merged into 'Other Unclassified (U)', given there are rarely records of C and M road classes in the cycling accident database. Consequently, ‘Combined road class’ in this paper has the following states: A-A, A-B, A-U, B-U, B-B, U-U and same road (means not at the junction).

9) Speed limit of a city road

When driving on road, the driver must not drive faster than the speed limit for the type of road and type of vehicle. According to the *Highway Code, road safety and vehicle rules* of the UK, a speed limit of 30mph the most widely applied compared to other limits. (https://www.gov.uk/speed-limits). ‘Speed Limit’ is therefore classified into three states: 30mph, above 30mph andbelow 30mph.

10) Road type

The road type associate with a cycling accident refers to the main carriageway on which the accident occurs. STATS19 lists six road types for cycling accidents: roundabout, one way street, dual carriageway, single carriageway, slip road, and unknown road. Nevertheless, among all the collected accident reports, very few occurred on one-way street, slip road and unknown road and none of them caused fatal consequence and hence these three road types are merged into one state – ‘others’. This variable has four states: roundabout, single carriageway, dual carriageway and others.

11) Junction detail

Junction is defined as a place where two or more roads meet with various angles of the axes of the roads. If there are two or more junctions within 20 meters of an accident, the junction that is the closest to the accident is recorded in the report. The UK government classifies junctions into nine categories. Some of them are not relevant in this paper since no relevant accident data are associated with them appropriately. Consequently, the processed states for ‘junction detail’ are crossroads, roundabout, not at junction, T or staggered junction and ‘other’ junction.

12) Junction control

The existence of control measures at a junction is crucial for reducing the severity of accidents because they are effective in regularizing the behaviour of road users. Different control measures have different efficiency. This variable has four states: automatic traffic signal, give way, ‘other’ control measures and no control.

13) Manoeuvre of cyclists

The manoeuvre of a cyclist refers to the action(s) taken immediately before the occurrence of an accident. According to STATS19 reports, there are 18 types of possible manoeuvres. Based on the availability of the relevant data and the actions having similar nature, five states are presented for this variable: go ahead, overtaking vehicles, turning, waiting at junction, and ‘others’.

14) Skidding, cyclist location and first point of impact (Cyclist)

These variables’ meanings are self-explanatory while their states are set as recommended by the STATS19 reports as follows.

Skidding – Yes, No

Cyclist location – Main carriageway, Not

First point of impact – Front, Back, Offside, Nearside, Not impact

15) Age of cyclists

Based on the information provided by United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the National Statistics Office of the UK, persons are divided into four groups according to their age bands within the cycling safety context: Child (Under 15), Youth (15-24), Working adult (24-65) and the elderly (Over 65).

16) Gender of cyclist

Two states for this variable are male and female.

17) Contributory factors (see Table 1).

All the contributory factors in the BN model has two states: Yes, No.

4.1.2 Influenced/Target variable

With regard to the influenced variable ‘accident severity’, a cycling accident can lead to a fatal, serious, or slight consequence (STATS19 report). Fatal accidents include only those cases where a death occurs in less than 30 days as a result of the accidents. Serious accidents contain the cases of broken necks or backs, severe head/chest injuries, internal injuries, loss of arms/legs, deep penetrating wounds, crushing, concussion, and others. Slight accidents refer to those accidents causing neck pain, bruising, sprains and strains, etc. The detailed information and the full list of injury for three severity types are found in a STATS19 report. Given the above description, Table 1 presents the full list of risk variables influencing the severity of cycling accidents in the BN, and their defined states.

Table 1. Identified risk factors and their states

VARIABLE	STATE
District	Knowsley, Sefton, Liverpool, St.Helens, Wirral
Day	Mon, Tues, Wed, Thur, Fri, Sat, Sun
Time	Rush hour, Non-Rush hour
Encountering vessel types	Cars, HGV, PSV, Motorcycle, Cyclist, Other/unknown
Weather	Fine with high winds, Fine without high winds, Rain with high winds, Rain without high winds, Other
Road surface	Dry, Wet/Damp, Other
Street lighting	Dark (no street light or unknown), Dark (street lights present and lit), Daytime
Combined road class	A-A, A-B, A-U, B-B, B-U, U-U, Same road
Speed limit	Below 30km/h, 30km/h, Above 30km/h
Road type	Single Carriageway, Dual carriageway, Roundabout, Other
Junction detail	Crossroads, T or staggered, Not at junction, Roundabout, Other
Junction control	Automatic traffic signal, Give way/Uncontrolled, Other way, None
Manoeuvre of Cyclist	Go ahead other, Overtake vehicles, Turning, Waiting at junction, Other
Skidding	Yes, No
Cyclist location	Main carriageway, Other
First point of impact	Front, Back, Offside, Nearside, Not impact
Age of cyclist	Child, Youth, Working adults, Elderly
Gender of cyclist	Male, Female
Accident severity	Slight, Severe, Fatal
Contributory factor of Victim	Yes, No
Contributory factor of Other road users	Yes, No

4.2 Structure of BN Model

Through the process of TAN learning, the structure of BN model for the analysis of cycling accident severity is developed and shown in Figure 2.

The ‘V’ nodes in the network represent the contributory factors caused by victim (cyclist), while ‘O’ nodes represent the contributory factors caused by the encountering road users in an accident. Each of ‘V’ and ‘O’ has 16 possible contributory factors as explained in Appendix 2, numbered from 1 to 16 (i.e. V1-V16 and O1-O16). The definition of each contributory factor is presented in Table A3.1.

In the process of data-driven structure learning in BN, the network structure is purely learnt from a mathematical perspective, which results in the existence of some links that only have statistical dependence but could not reflect the reality and needs to use expert knowledge to conduct adjustment for a fine-tuned network (Wang & Yang, 2018). Through investigating the opinions and judgments from domain experts (i.e. police officer, staff in transport department and cycling communities), as well as the prior knowledge learnt from the previous work to verify the initial network generated purely by data, the following links and arcs in the initial TAN model are not consistent with the actual situation and hence removed.

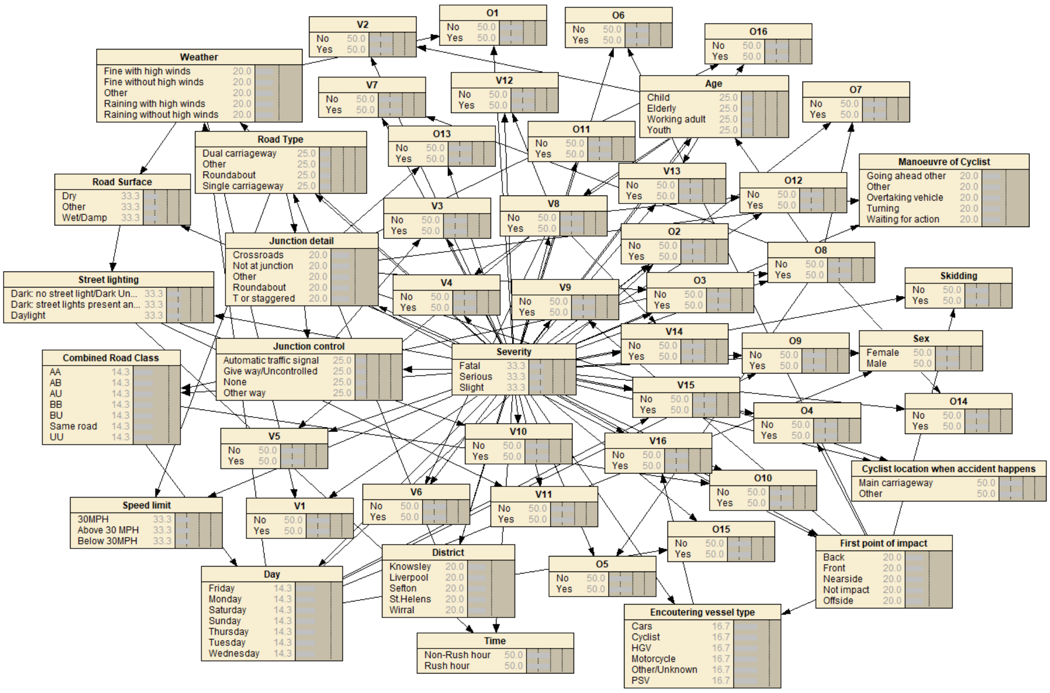

V2
O1
O6
O16
Weather: Fine with high winds, Fine without high winds, Other, Raining with high winds, Raining without high winds
V7
V12
Age: Child, Elderly, Working adult, Youth
O7
Road Type: Dual carriageway, Other, Roundabout, Single carriageway
O13
O11
Road Surface: Dry, Other, Wet/Damp
V13
O12
Manoeuvre of Cyclist: Going ahead other, Other, Overtaking vehicle, Turning, Waiting for action
V3
V8
O2
O8
Junction detail: Crossroads, Not at junction, Other, Roundabout, T or staggered
Street lighting: Dark: no street light/Dark Un..., Dark: street lights present an..., Daylight
V4
V9
O3
Skidding
V14
O9
Sex: Female, Male
Combined Road Class: AA, AB, AU, BB, BU, Same road, UU
Junction control: Automatic traffic signal, Give way/Uncontrolled, None, Other way
Severity: Fatal, Serious, Slight
V15
O4
O14
V5
V10
V16
Cyclist location when accident happens: Main carriageway, Other
O10
Speed limit: 30MPH, Above 30 MPH, Below 30MPH
V1
V6
V11
O15
First point of impact: Back, Front, Nearside, Not impact, Offside
Day: Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday
District: Knowsley, Liverpool, Sefton, St.Helens, Wirral
O5
Encountering vessel type: Cars, Cyclist, HGV, Motorcycle, Other/Unknown, PSV
Time: Non-Rush hour, Rush hour
No, Yes

Figure 2. TAN structure for cycling accident severity

- ‘Day-Weather’, ‘Road Surface-Street lighting’, ‘Street lighting-Time’, ‘Street lighting-V10’, ‘Combined road class-Day’, ‘Day-Sex’, ‘Day-O9’, ‘Day-O2’, ‘Day-V15’, ‘First point of impact-Skidding’, ‘Age-V2’, ‘Road type-District’, ‘Junction detail-O13’, ‘Combined road class-O10’, ‘O12-V8’, and ‘Encountering vessel type-First pint of impact’.

Consequently, after the adjustment, the fine-tuned structure (based on Figure 2) is improved and presented in Figure 3.

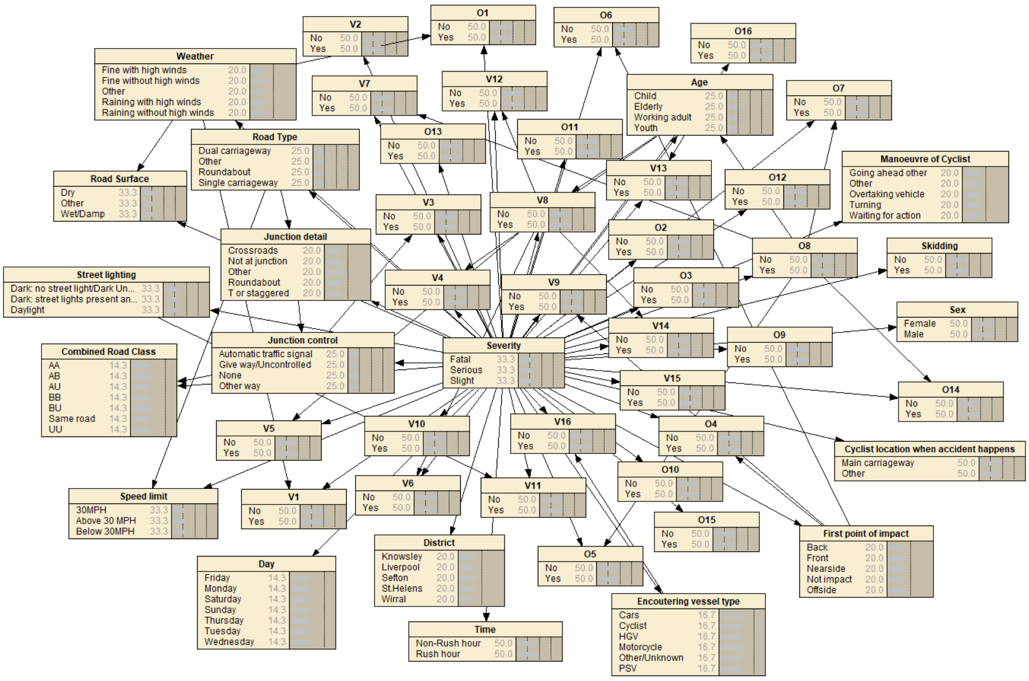

V2
O1
O6
O16
Weather: Fine with high winds, Fine without high winds, Other, Raining with high winds, Raining without high winds
V7
V12
Age: Child, Elderly, Working adult, Youth
O7
Road Type: Dual carriageway, Other, Roundabout, Single carriageway
O13
O11
Road Surface: Dry, Other, Wet/Damp
V13
O12
Manoeuvre of Cyclist: Going ahead other, Other, Overtaking vehicle, Turning, Waiting for action
V3
V8
O2
O8
Skidding
Junction detail: Crossroads, Not at junction, Other, Roundabout, T or staggered
Street lighting: Dark: no street light/Dark Un..., Dark: street lights present an..., Daylight
V4
V9
O3
Sex: Female, Male
Combined Road Class: AA, AB, AU, BB, BU, Same road, UU
Junction control: Automatic traffic signal, Give way/Uncontrolled, None, Other way
Severity: Fatal, Serious, Slight
V14
O9
V15
O14
V5
V10
V16
O4
Cyclist location when accident happens: Main carriageway, Other
O10
Speed limit: 30MPH, Above 30 MPH, Below 30MPH
V1
V6
V11
O15
First point of impact: Back, Front, Nearside, Not impact, Offside
Day: Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday
District: Knowsley, Liverpool, Sefton, St.Helens, Wirral
O5
Encountering vessel type: Cars, Cyclist, HGV, Motorcycle, Other/Unknown, PSV
Time: Non-Rush hour, Rush hour
No, Yes

Figure 3. Fine-tuned BN structure for cycling accident severity

4.3 Model results

After the structure of the BN, the conditional probabilities of the involved nodes are obtained via a gradient descent approach (Yang et al., 2018b). Based on the calculated conditional probability table (CPT) of each node, the result of the TAN model is presented in Figure 4. It indicates that the probability of a slight accident is 76.3%, a serious accident is 23.2% and a fatal accident is 0.48%. When calculating these values directly from the accident report database, it was found that the similarity between the two are very high. For instance, the direct statistical analysis reveals that the probability of having a slight accident is 75.76% (by dividing the total number of slight accidents by the total number of the accident reports), the probability of serious accident 23.75%, and the probability of fatal accident 0.49%. This proves the accuracy of the results that the model delivers based on the historical data.

Based on the proposed model, the unobserved situations associated with the cycling accidents can be predicted through the generated posterior probabilities when observable evidence is provided. Therefore, the BN model is served as a dynamic prediction tool to foreseen the accident severity degree under different situations. Before that, the model prediction ability is tested using newly collected accident data collected from 2018.

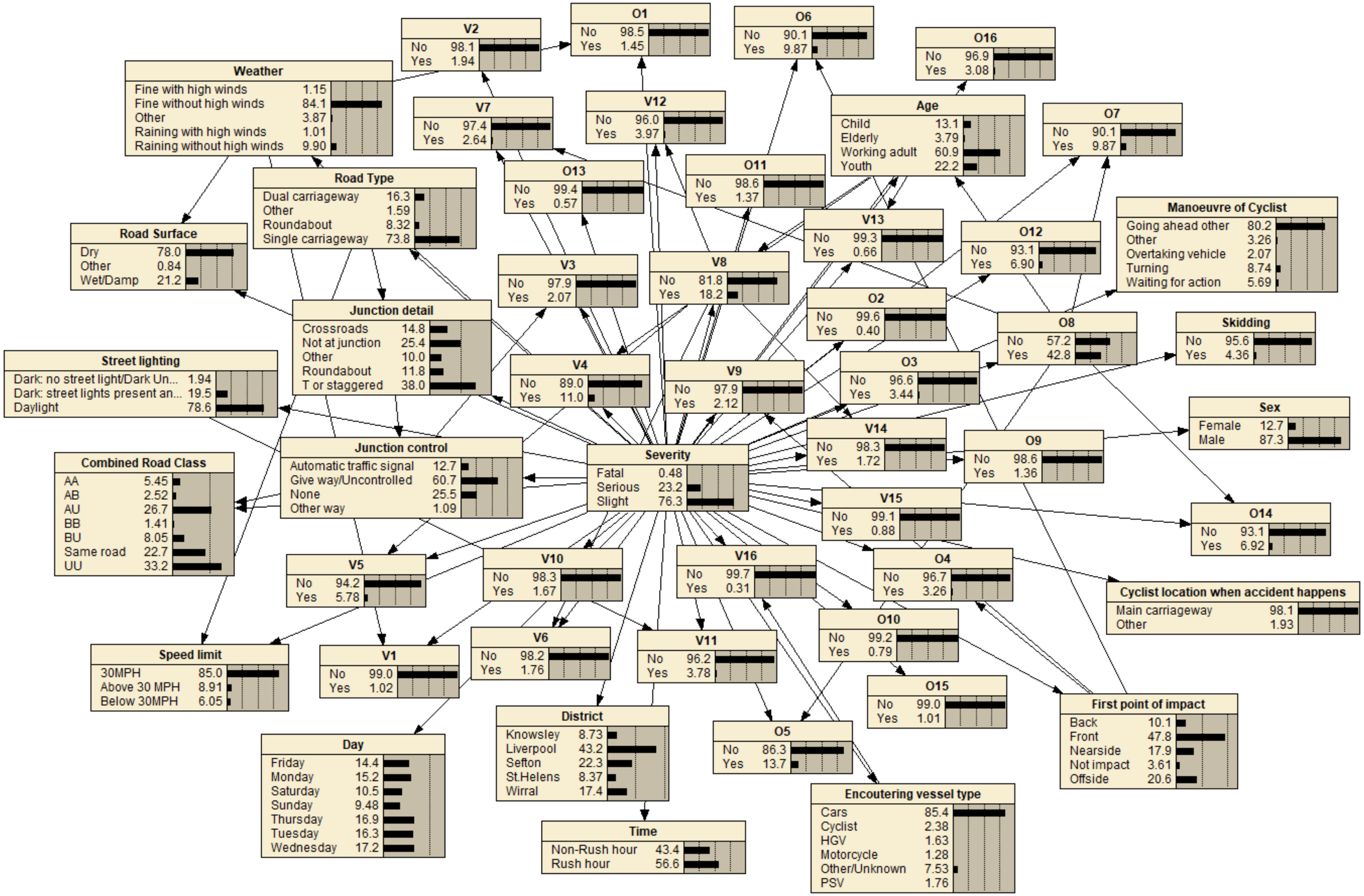


Figure 4. BN model result (marginal probability) based on 2258 historical reports

4.4 Model verification

The verification process focuses on two aspects: one is the prediction performance of the model, the other is consistency test of the model. To verify the proposed model, 235 new accident cases in the Liverpool region within 2018 were collected from Merseyside Police. Among these reports, a few of them contains incomplete important information related to the defined risk variables (in Table 1), i.e. contributory factors, cyclist gender, and manoeuvre details. They are excluded and as a result, 213 accident reports are finally used for model verification.

4.4.1 Prediction performance

Relevant information of the 213 cycling accidents is used individually to test the proposed model (Figure 4), the state of accident severity with the highest probability is used as the result delivered by the proposed model. The following table reveals the accuracy rate of the BN risk model in predicting different severity types of cycling accidents by comparing the model results with the ones in real accident reports.

Table 2. Model accuracy

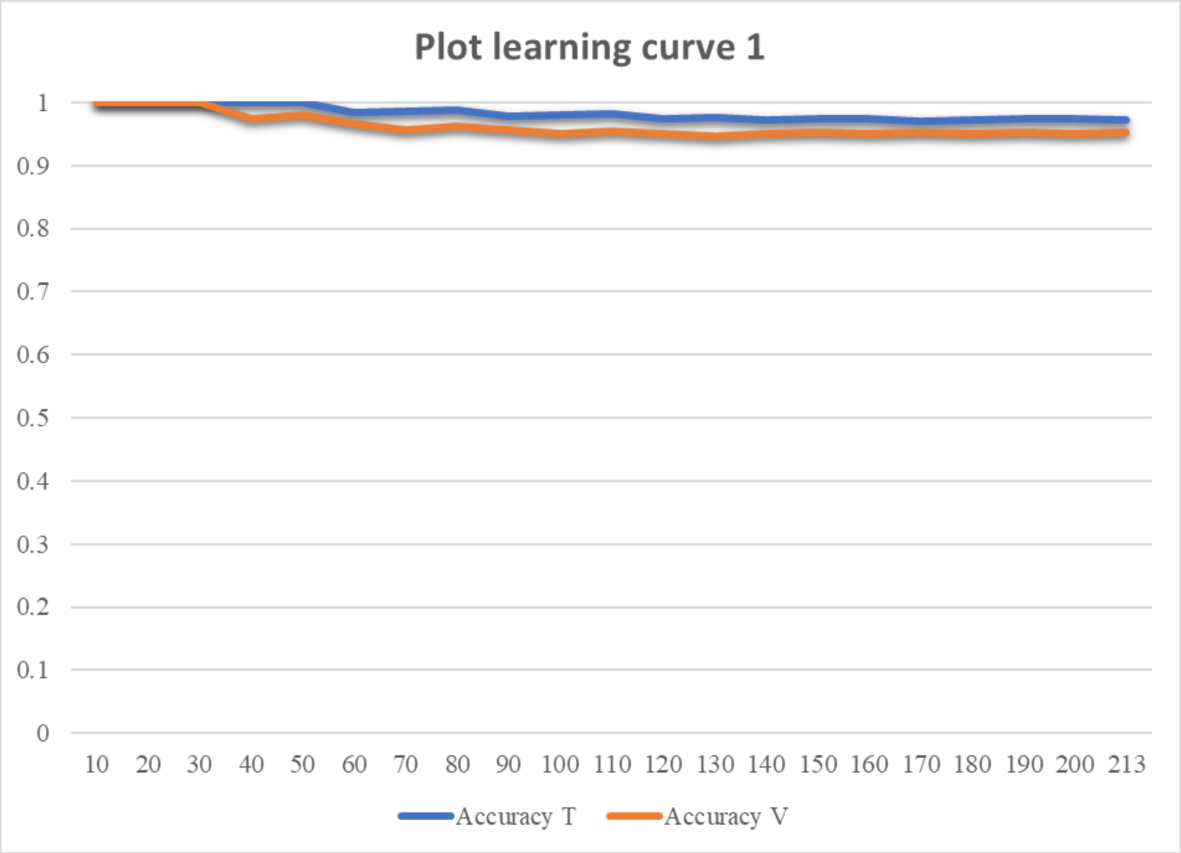
Model delivery **Real severity**	**Slight**	**Serious**	**Fatal**	**Total number**	**Accuracy**
Slight	**168**	8	0	176	95.45%
Serious	2	**34**	0	36	94.44%
Fatal	0	0	**1**	1	100%
General	170	42	1	213	95.31%

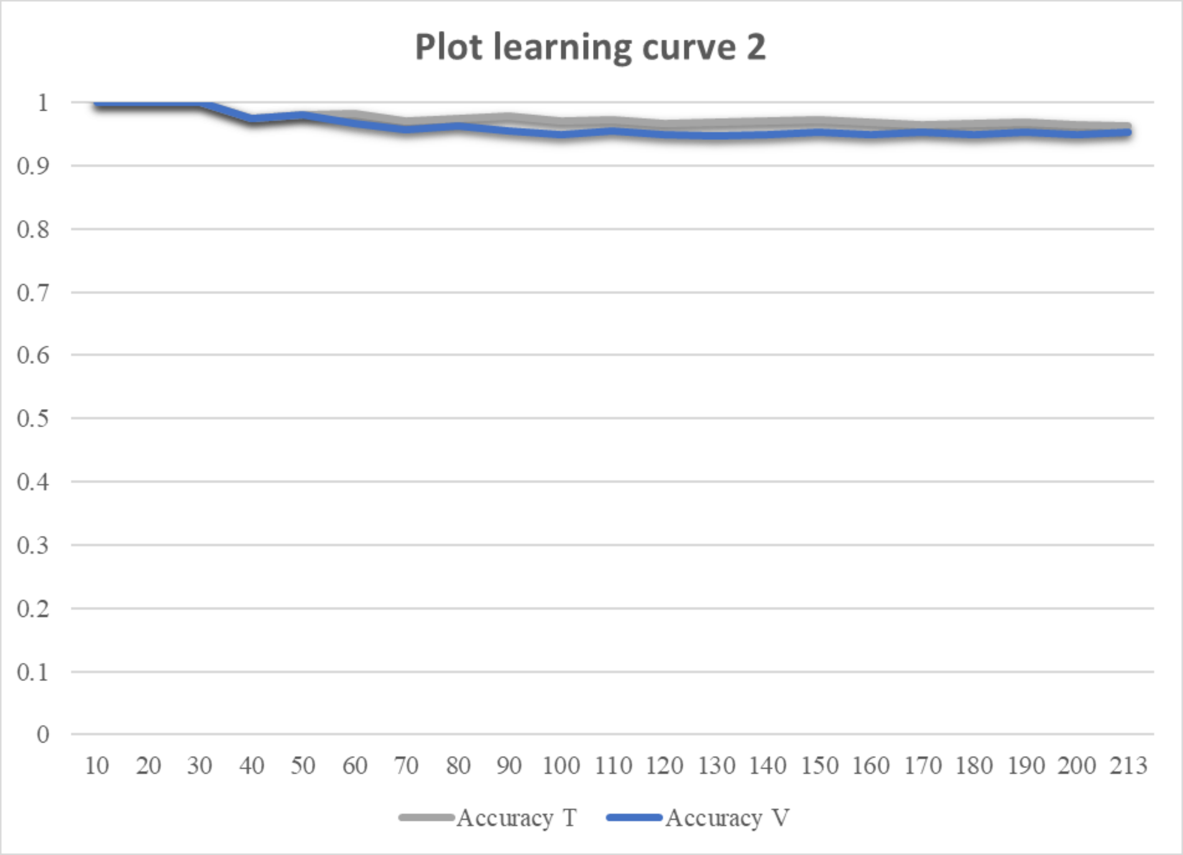
To explain Table 2, an example of ‘slight severity’ is used. Among the 213 cycling accidents, 176 have a slight severity consequence. When incorporating the information (against each of the risk variables in Table 1) of each cycling accident into proposed model, in 168 accidents the model suggests a slight consequence, while 8 receive a serious consequence. Therefore, the accuracy rate for ‘slight severity’ is calculated as 95.45% (168/176). The same goes to ‘serious severity’ and ‘fatal severity’. From Table 2, the accuracy rates of ‘slight severity’, ‘serious severity’ and ‘fatal severity’ is 95.45%, 94.44% and 100% respectively, indicating the model is reliable in terms of providing accurate and consistent forecasting results. Additionally, its overall accuracy rate is 95.31% (203/213).

Furthermore, when we calculate the training accuracy of the proposed model, it is 97.18% (2205/2269). The similarity between training accuracy and validation accuracy indicates our model shows a good fit. Specifically, a plot learning curve could be resorted to consolidate this conclusion.

A learning curve is a plot of model learning performance over experience or time. Learning curves (LCs) are deemed effective tools for monitoring the performance of models exposed to new evidence. It is widely used in machine learning for algorithms that learn (optimize their internal parameters) incrementally over time (Michel et al., 2011). A LC consists of two parts: one is a train learning curve, the other is a validation learning curve. The x-axis in the curve represents number of sample inputs, while the y-axis in the curve represents the accuracy of the model.

In this research, to draw the learning curve of the proposed model, the same number of samples in training data is randomly selected to match the validation data. The process is repeated for three times, and the learning curves are presented in Figure 5 as follows.


Plot learning curve 1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 213
Accuracy T
Accuracy V


Plot learning curve 2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 213
Accuracy T
Accuracy V

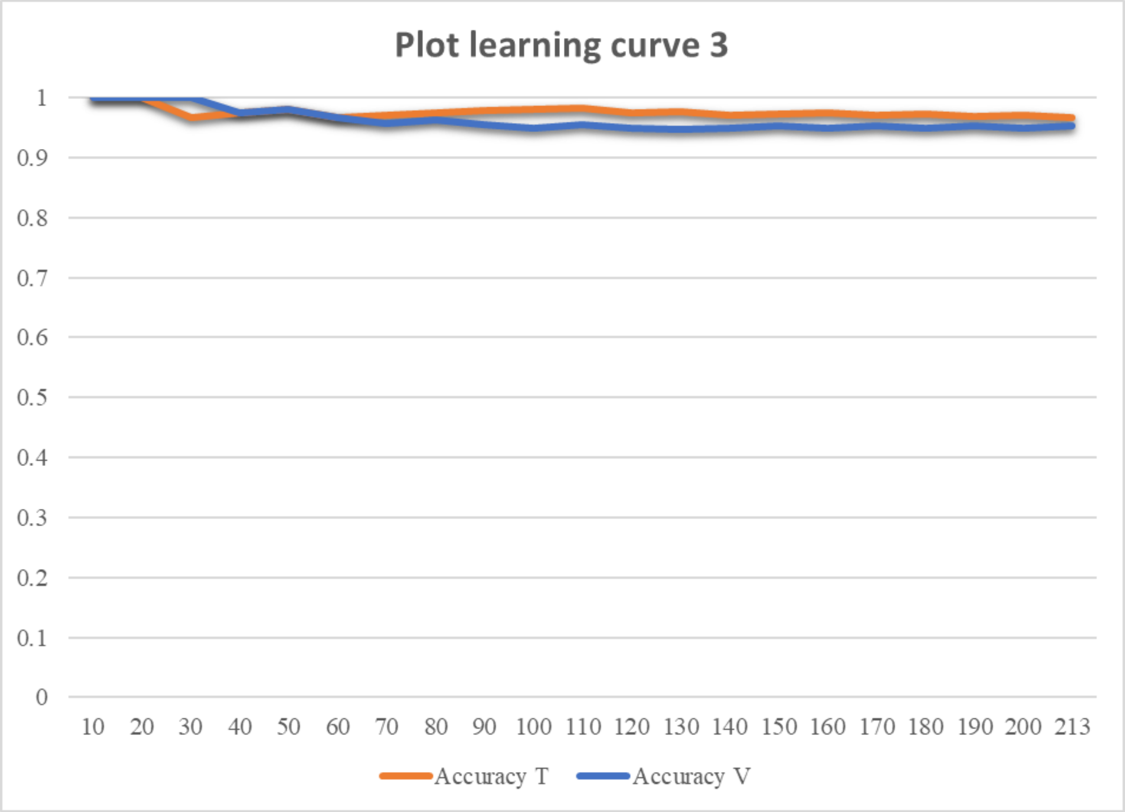


Figure 5 Plot learning curves of the proposed model

It is observed from Figure 5 that the model is of good fit and does not exist overfitting or underfitting problems because of

- The curves maintain at a point of stability with the sample number increases.
- There is only a subtle gap between two curves, which means the training accuracy and the validation accuracy is consistent.

Therefore, the proposed cycling model could be used as a reliable prediction tool.

4.4.2 Kappa statistic for model consistency test

In this research, the consequence severity levels are unbalanced with the majority being slight injuries. In this case, using the percent calculation along for the model accuracy prediction and validation are arguably insufficient. Kappa statistic, as an alternative statistical approach, is used to test the model consistency. Since there are two raters in this research (predicted results and real results), Cohen's kappa coefficient is selected for the model validation.

The calculation process is shown as follows:

$$p_e = \frac{1 \times 1 + 42 \times 36 + 170 \times 176}{213 \times 213} = 0.6897, \qquad p_0 = 0.9531$$

$$\boldsymbol{k} = \frac{0.9531 - 0.6897}{1 - 0.6897} = 0.8488$$

The Cohen's kappa ($\boldsymbol{k}$) is 0.8488. Based on the guidelines from Altman (1999), a kappa ($\boldsymbol{k}$) of 0.8488 represents a strong strength of agreement, which means the model is strongly consistent with the real accident consequences.

The detailed information and calculations for the comparison between real consequences and the model predicted results are found in Appendix 3.

5. Sensitivity analysis

5.1 Mutual information analysis

According to Equation (2), the mutual information between ‘accident severity’ and other risk variables is obtained and demonstrated in Table 3. The ‘percentage’ column in the table represents the extent to which the shared information between the ‘accident severity’ and other variables belongs. In other words, the percentage value of each variable indicates its individual impact on ‘accident severity’. The mutual information values in the table are independent and irrelevant to others.

Table 3. Mutual information of ‘Accident severity’ and other variables

Variable/Node	Mutual Info	Percentage
Age	0.01122	1.36
District	0.00818	0.995
Day	0.00778	0.945
Encountering vessel type	0.00654	0.795
First point of impact	0.00595	0.724
Combined Road Class	0.00438	0.532
Junction control	0.00406	0.493
Junction detail	0.00376	0.457
Maneuver of Cyclist	0.00354	0.43
V4	0.00234	0.284
V8	0.00224	0.2727
O4	0.00189	0.229
Road Type	0.00188	0.229
V9	0.0017	0.207
Speed limit	0.00166	0.202
Weather	0.00164	0.199
V5	0.0016	0.194
V10	0.00143	0.174
V12	0.00142	0.173
V3	0.0014	0.171
O16	0.00132	0.161
O10	0.00126	0.154
O9	0.0012	0.146
V1	0.00114	0.138
V7	0.00087	0.106
V11	0.00075	0.091
O14	0.00073	0.0887
Cyclist location when accident happens	0.00067	0.08110
Street lighting	0.00066	0.0804
O5	0.00054	0.0655
O3	0.00045	0.0548
O11	0.00041	0.0495
Time	0.00038	0.046
O8	0.00038	0.0458
Road Surface	0.00029	0.0357
V13	0.00029	0.0354
V2	0.00025	0.0307
V6	0.00025	0.0298
O7	0.00022	0.0273
O15	0.0002	0.0245
V14	0.00018	0.0216
Skidding	0.00017	0.0205
O13	0.00015	0.0183

O1	0.00011	0.013
O12	0.00009	0.0114
V16	0.00005	0.00662
Sex	0.00005	0.00577
V15	0.00005	0.00563
O6	0.00001	0.00149
O2	0.00001	0.0013

To understand the role of variables more clearly, data clustering is necessary. As a statistical classification technique for discovering whether the individuals of a population fall into different groups (Merriam-Webster Online Dictionary, 2008), data clustering is useful for our research to gain insight into risk variables, as well as identify the degree of similarity among them. As a powerful approach, the K-means clustering algorithm is selected for data clustering.

It is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters. In the clusters, each observation belongs to the cluster with the nearest mean (cluster centres or cluster centroid). In other words, to find a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized (Anil, 2010). Although the K-means clustering algorithm was first proposed over 50 years ago, it is still one of the most widely used clustering algorithms because of its easiness of implementation, simplicity, efficiency and empirical success. The goal of K-means clustering algorithm in our research is to discover the natural grouping of the influencing variables in the model.

Following the main steps of this algorithm are as follows (Jain & Dubes, 1988):

1. Select an initial partition with K clusters; repeat steps 2 and 3 until the cluster membership stabilizes.

2. Generate a new partition by assigning each pattern to its closest cluster centre.

3. Compute new cluster centres.

In this research, the number of clusters (*k*) is three according to the Elbow method. The idea of the Elbow method is to run k-means clustering on the dataset for a range of values of k (say, k from 1 to 10 in our research), and for each value of k to calculate the sum of squared errors (SSE). Then, the next step is to plot a line chart of the SSE for each value of k. The elbow of the curve is the value of k that is the best (Robert, 1953). The results of K-means clustering are shown in Table 4.

Table 4. K-means clustering results

Class	1	2	3
Objects	5	18	27
Sum of weights	5	18	27
Within-class variance	0.061	0.016	0.001
Minimum distance to centroid	0.019	0.011	0.004
Average distance to centroid	0.171	0.100	0.028
Maximum distance to centroid	0.396	0.270	0.096

Once the *k* is determined, the variables can be classified based on the results obtained from K-means clustering results. Specifically, the variable assignment is based on the ‘sum of weight’ information in the table. For example, 'sum of weight' of 'class 1' is 5, which means there should be 5 variables in class 1. Hence, there are 5 variables in class 1, 18 variables in class 2, and 27 variables in class 3.

Next step is to assign the variables to the three defined classes according to the mutual information. The results of k-means clustering suggest the influencing degree from class 1 to class 3 is decreasing, indicating the variables having top 5 mutual information values should be assigned to class 1, which is also named as first priority variables. The same mechanism is used to assign the variables to the classification of class 2 (2nd priority variable) and class 3 (low priority variable).

Table 5 presents the variables included in each class from the K-means clustering results, which is also denoted as '1st priority variables', '2nd priority variables', and 'low priority variables'.

Table 5. Variable classification

Class 1 (1st priority variables)	Age, District, Day, Encountering vessel type, First point of impact
Class 2 (2nd priority variables)	Combined Road Class, Junction control, Junction detail, Manoeuvre of Cyclist, V4, V8, O4, Road Type, V9, Speed limit, Weather, V5, V10, V12, V3, O16, O10, O9
Class 3 (low priority variables)	V1, V7, V11, O14, Cyclist location when accident happens, Street lighting, O5, O3, O11, Time, O8, Road Surface, V13, V2, V6, O7, O15, V14, Skidding, O13, O1, O12, V16, Sex, V15, O6, O2

The variable classification provides important insights for the risk informed policy making and implementation from both cyclists and transport authorities perspectives. These priorities are listed here:

1) 1st priority variables have the highest-level influence on the accident severity. Transport authorities should pay the highest attention when formulating relevant policies, as well as inform cyclists for their implementation and accident prevention. Financial resources should be allocated to address them with the highest priority.

2) 2nd priority variables are less influential compared to those from the first category. However, they will also threaten the cycling safety and the sequence of addressing such factors can refer to the cost benefit analysis of their associated risk control measures and policies.

3) Low priority variables have little influence on cyclist safety. Their risk contribution should be more considered when combined with the factors from the first two categories in Section 5.2.

5.2 Influence of multiple variables

In this section, the influence of multiple variables in a combined way on 'accident severity' is evaluated. As shown in Section 5.1, the influence of single variables on 'accident severity' is not obvious because the values of mutual information are relatively small. However, the occurrence of a cycling accident normally results from the simultaneous presentation of multiple risk variables, which means some specific combination of risk variables will generate a much greater impact on 'accident severity' compared to the simple numerical aggregation of their individual influence values. Additionally, from a policy making perspective, it may sometimes be financially infeasible to control some high-prioritised influence variables, or impossible to control them (i.e. age, district). On this occasion, understanding how the combined sets involving the high-prioritised variables which leads to major or fatal consequence can help transport authorities make realistic and effective safety policies.

As there are 50 variables in the model, it is too complicated to enumerate all the combinations, hence in this paper the combinations of two variables is selected to test the influence of multiple risk variables on accident severity. Based on the classification of variables in the last section, the analysis on the two-variable combinations is conducted via the following aspects: '1st priority-1st priority' combination, '1st priority-2nd priority' combination, '1st priority-low priority' combination, '2nd priority-2nd priority' combination, '2nd priority-low priority' combination' and 'low priority-low priority' combination.

It is noteworthy that only the worst case is considered when analysing the influence brought by each category, as the worst case often provides the most valuable information for safety policy making.

1) 1st priority variable – 1st priority variable

Two variables selected in this section are 'cyclist age' and 'encountering vessel type'. Table 6 illustrates the occurrence probability of different accident severity when the worst case happens.

Table 6. Combination of 1st and 1st priority variables

Cases	Cyclist age	EVT	Fatal	Serious	Slight
General case	/	/	0.48	23.1	76.4
Worst-Cyclist age	Elderly	/	3.49 (+627%)	44.1(+91%)	52.5
Worst-EVT	/	Motorcycle	6.89 (+1335%)	37.9 (+64%)	55.2
Worst combination	Elderly	Motorcycle	30.2 (+6192%)	45.8 (+98%)	24.1

(The number in the brackets is the rate of change, '+' means increase, '-'means decrease)

2) 1st priority variable – 2nd priority variable

Two variables selected are 'cyclist age' and 'road type'. The estimated results delivered by the BN are presented in Table 7.

Table 7. Combination of 1st and 2nd priority variables

Cases	Cyclist age	Road type	Fatal	Serious	Slight
General case	/	/	0.48	23.1	76.4
Worst-Cyclist age	Elderly	/	3.49 (+627%)	44.1(+91%)	52.5
Worst-Road type	/	Dual carriageway	1.08 (+125%)	25.3 (+10%)	73.6
Worst combination	Elderly	Dual carriageway	7.21 (+1402%)	45.2 (+96%)	47.6

3) 1st priority variable – low priority variable

'Cyclist location when accident happens' and 'Encountering vessel type' are chosen as the target variables, and the results are presented in Table 8.

Table 8. Combination of 1st and insignificant priority variables

Cases	EVT	Cyclist location	Fatal	Serious	Slight
General case	/	/	0.48	23.1	76.4
Worst-EVT	Motorcycle	/	6.89 (+1335%)	37.9 (+64%)	55.2
Worst-Cyclist location	/	Other	2.28 (+375%)	27.2 (+18%)	73.6
Worst combination	HGV	Other	13.7 (+2754%)	40 (+73%)	46.3

4) 2nd priority variable – 2nd priority variable

Selected variables are 'Road type' and 'V4' and the result is shown in Table 9.

Table 9 Combination of 2nd and 2nd priority variables

Cases	Road type	V4	Fatal	Serious	Slight
General case	/	/	0.48	23.1	76.4
Worst-Road type	Dual carriageway	/	1.08 (+125%)	25.3 (+10%)	73.6
Worst-V4	/	Yes	1.20 (+150%)	28.8 (+24.7%)	70.0
Worst combination	Dual carriageway	Yes	2.56 (+433%)	31.2 (+35.1%)	66.2

5) 2nd priority variable – low priority variable

Selected variables are 'Road type' and 'Cyclist location when accident happens' and the result is shown in Table 10.

Table 10. Combination of 2nd and low priority variables

Cases	Road type	Cyclist location	Fatal	Serious	Slight
General case	/	/	0.48	23.1	76.4
Worst-Road type	Dual carriageway	/	1.08 (+125%)	25.3 (+10%)	73.6
Worst-Cyclist location	/	Other	2.28 (+375%)	27.2 (+18%)	73.6
Worst combination	Dual carriageway	Other	6.28 (+413%)	25.3 (-25%)	68.4

6) Low priority variable – low priority variable

Selected variables are 'Sex' and 'Cyclist location when accident happens' and the result is shown in Table 11.

Table 11. Combination of low priority and low priority variables

Cases	Sex	Cyclist location	Fatal	Serious	Slight
General case	/	/	0.48	23.1	76.4
Worst-Sex	Male	/	0.5 (+4%)	23.1 (/)	76.4
Worst-Cyclist location	/	Other	2.28 (+375%)	27.2 (+18%)	73.6
Worst combination	Dual carriageway	Other	2.18 (+354%)	21.2 (-8%)	70.6

It is evident that from the analysis in Tables 6-11 that the following remarks can be made for possible implications in Section 5.3:

- It is obvious to find that the combination of risk variables has much greater influence on the accident severity than individually, regardless of the variable categories.
- The higher priorities the risk variables have, the greater influence they will have on accident severity when grouping together.
- According to the probabilities of fatal and serious consequence in these tables, the combination of two '1st priority variables' can increase the likelihood of fatal consequence for dozens of times, which is more life threatening than other types of combinations for cyclists.

- When grouping together with specific 1st priority variables, an insignificant variable can also play an important role in affecting accident severity, even double the likelihood of fatal consequence of that from 1st priority variable individually.
- Compared with fatal consequence, the change rates of serious consequence are not remarkable under various situations.

5.3 Discussion on the outcomes – Features in Liverpool city region

The following results from the model reveal the new findings of useful insights for improving cycling safety in the case city region and related suggestions derived from previous studies for risk reduction. The findings of this papers have been separated into these two sections accordingly.

5.3.1 Findings

In this subsection, the findings derived from the model are explained in detail with regard to the analysis and corresponding suggestions.

1) Collisions with motorcycles are the most dangerous situation for cyclists, which has much higher probability of causing a fatal consequence and serious consequence than encountering with other road users, i.e. cars, PSV, or HGV, according to Table 12. For example, the probability of being caught in a fatal/serious consequence with a motorcycle is 2.6times/1.4times respectively higher than encountering with HGV, which is the second highest vehicle types.

Table 12. Probability encountering different vessel types

Situation	Probability (Fatal/Serious/Slight)
General case	0.48/23.2/76.3
Encountering Cars	0.26/22/7/77.1
Encountering Cyclists	1.85/22.2/75.9
Encountering HGV	2.7/27/70.3
Encountering PSV	2.5/20/77.5
Encountering Motorcycle	6.91/37.9/55.2
Encountering Other	0.59/26.9/72.5

Based on a careful analysis on the collected accident data, motorcyclists in the Liverpool region are more likely to be in a state of emotional riding, for example, riding in an aggressive or dangerous manner, behaved in a negligent or thoughtless manner, or in a hurry mode. Through the investigation and interview on these motorcyclists according to the accident report, it shows that lack of concern about the possible consequences of their actions (careless), acts in spite of the likely consequences (reckless), or fails to consider the consequences of their actions as a result of being in a hurry are the main reasons for their behaviours, which eventually leads to serious consequences.

On the other hand, in many cases, the speed of motorcycles is as high as other motor vehicles (i.e. cars, HGV, PSV). However, motorcyclists are more vulnerable to crashes and accidents as they are exposed and have limited safety equipment available to protect them as compared with other vehicle users, which have seat belts, airbags and other safety features – this also applies to cyclists. Therefore, the consequences brought to both cyclists and motorcyclists if an accident happened between them would be more serious.

Several suggestions could be put forward to help reduce the risk in this perspective:

- Bicycle-related equipment

Use of rear or pedal reflectors; Improve bicycle security; Cyclists need to wear reflective clothing; Visibility aids (fluorescent vests, flashing lights on clothing, fluorescent clothes, reflectors or reflective strips); Promotion of the use of safety gears (bicycle lights, reflector, helmet).

- Traffic rules & policy regulation

Formulating traffic laws which give special consideration to vulnerable road users, especially cyclists.

2) The study found that injury severity of cyclists involved in traffic crashes increased with road speed limits (see Table 13).

Table 13 Probability under different speed limit

Situation	Probability (Fatal/Serious/Slight)
Above 30 MPH	1.00/27.7/71.3
30MPH	0.73/27.07/7/72.20
Below 30MPH	0.40/22.50/77.10

It is not surprised to reach this conclusion. As cyclists are vulnerable to risks compared with other road users, high speed limit roads will amplify their exposure to high cycling risks. Previous studies also found high speed limits could increase the risk of cyclist crashes as well as serious injury and fatality (Bíl et al., 2010).

Possible countermeasures with regards to this finding from the implication perspectives include:

- Separating bicyclists from high-speed traffic, for example separate bicycle paths on roadways that have a high-speed limit
- Low speed limit in residential neighbourhoods with significant bicycle traffic (Kim et al., 2007).

3) The elderly and the child are more likely to be severely injured as a result of a crash than the other age group.

Table 14 Probability of different age group

Situation	Probability (Fatal/Serious/Slight)
Elderly	3.49/44.21/52.3
Child	0.67/24.03/7/75.30
Working adult	0.36/20.13/79.51
Youth	0.20/19.30/80.50

It is found the injury severity of the elderly is much higher than the other age groups as seen in Table 14. Meanwhile, child is also more likely to be involved in severe accidents compared with working adults and the youth. This is because older cyclists are particularly vulnerable to severe injury mainly due their overall physical and mental fragility (i.e., slow reaction to unexpected situations, little cycling experience and physical illness) often compounded by pre-existing conditions (Anstey et al., 2005).

Suggestions for possible improvements in terms of this finding include:

- Traffic laws which give special consideration to the vulnerable road users.

- More attention paid to youngest (under 20) and oldest cyclists (over 65) when formulating traffic rules.

- Training education for all road users.

- Promotion of the use of safety gears (bicycle lights, reflector, helmet).

4) When a cycling accident happens at a junction in the Liverpool city region, the probability of having a fatal consequence is much lower than the one relating to the other locations, which are demonstrated in Table 15.

Table 15. Probability at different junction types

Situation	Probability (Fatal/Serious/Slight)
Crossroads	0.3/27.2/72.5
Roundabouts	0.002/22.8/7/77.198
T or staggered junction	0.2/22.5/77.3
Not at junction	1.04/23.46/75.5

The relevant research suggests that at junction, cyclists would be more careful and their safety awareness is high. Through an investigation at some major junctions in the Liverpool city region for a certain period, it was found that many cyclists will reduce their speed when approaching junctions to ensure there are no other vehicles driving towards/passing through them. Meanwhile, normally there will be different traffic control facilities at junctions, for example: 1) a police officer, traffic warden in uniform or school crossing patrol who is in control of the traffic; 2) automatic traffic signal; 3) stop sign; and 4) give way. These facilities will help protect the safety of cyclists and reduce the probability of being caught in some severe accidents. No traffic control will make the case worse, which is proved by our model as shown in Table 16.

Table 16. Probability of different junction control measures

Situation	Probability (Fatal/Serious/Slight)
Automatic traffic signal	0.69/28.4/70.9
Other (authorized person/stop sign)	0.01/12.1/7/87.8
Give way	0.22/22.3/77.5
None	1.1/23.4/75.5

The suggestions to improve this perspective should focus on the traffic control measures, including: separation of different user types in the areas with high traffic volume; improvement of road signalling and repression of traffic law infringement through more intensive policing.

5) Road surface condition has little impact on safe cycling in Liverpool. In fact, the research finding shows dry road surface even leads to more fatal accidents than that on a wet/damp road surface condition.

Table 17. Probability of male and female

Situation	Probability (Fatal/Serious/Slight)
Dray surface condition	0.6/23.4/76
Wet surface condition	0.4/22.8/7/76.8

The explanation is that in rainy days the speed of both cyclists and vehicles in the case area are found much slower. Among 1762 cases happened on dry surface condition reported in Liverpool, exceeding speed limitation is one of the major reasons, occupying 14.9%. On the other side, the number of over speeding in wet road is only 8.3%. This phenomenon well reflects the mentality of drivers and cyclists under different environments. When driving/riding on wet road surfaces, the road users are more gingerly and do not act recklessly.

Previous studies are also supportive to this finding. Rämä (2001) stated that the average speeds on a slippery road surface is lower than that in a good road surface conditions which are roughly 14 km/hour.

To reduce the effect of road surface condition on cycling safety, several suggestions related to road infrastructure are put forward: scheduled maintenance of bicycle-related infrastructures; keep cycling surfaces clean and decrease the number of obstacles on bicycle infrastructure.

6) Liverpool female cyclists have a lower chance of being caught in fatal/serious cycling accidents than males . From Table 18, in Liverpool, the probability of female cyclists being caught in a fatal accident is 70% less than of a male cyclist. In addition, the probability of a female cyclist being involved in a serious accident is 17.5%, which is 6% lower than a male cyclist.

Table 18. Probability of male and female

Situation	Probability (Fatal/Serious/Slight)
Male cyclists	0.5/23.2/76.3
Female cyclists	0.15/17.5/7/82.35

In fact, it is not surprising to find that the chance of female cyclists being caught in severe conditions is lower than male cyclists. Similar to the car insurance charge, male cyclists are more likely to be caught in dangerous situations due to the fact that:

- Male cyclists tend to cycle faster and are less likely to wear safe equipment
- A large proportion of male cyclists are young cyclists, which means they have a relatively low safety consciousness
- Male cyclists are more likely to participate in reckless activities than female, according to the accident reports
- The statistics also shows that male cyclists are more likely to ride after drinking, hence being more likely to be involved in fatal crashes

Suggestions to reduce this type of risk include:

- Publicity & Education

Training education for all road users; Improving poor cyclist behaviours by education and enforcement

- Promotion of the use of safety gears (bicycle lights, reflector, helmet)

7) In terms of the contributory factors, their occurrence from the encountering road user perspective will usually produce more dangerous consequences than that from the cyclists.

This is because cyclists are more vulnerable in accidents compared with the encountering road users in normal cases. According to the accident data collected in this research, almost all the victims (2250 of 2269, 99.2%) in these accidents are cyclists, hence more safety attention should be paid to more effectively engage other road users in cycling safety policy making.

Statistically, the contributory factors requiring extra attention are O16 (Other factors of other road users), O9 (Other road users’ reaction to unexpected cases), O4 (Injudicious action of other road users on road), V12 (Emotional riding of cyclist), and V10 (Physical/mental illness or impairment of cyclist), as these five factors are top 5 factors that put cyclists under severe situations than other contributory factors:

O16 includes special situations such as Stolen vehicles, Vehicles in course of crime, Vehicle door opened or closed negligently.

O9 includes sudden brake, swerved and loss of control of the encountering vehicles.

O4 includes injudicious actions of encountering road users like illegal turn, exceeding speed limit, following too close.

V12 includes emotional riding of cyclist such as aggressive riding, careless/reckless/in a hurry, and nervous/panic of the cyclists

V10 includes physical/mental illness of cyclists like impaired by alcohol/drugs/illicit medicine, fatigue, visual deficiencies and disability of the cyclists.

5.3.2 Findings to be further explored

Despite the fact that many findings from this study provide useful insights to guide the transport authorities to develop rational safety policies, there are still some findings requiring further exploration in future. They include

1) Cycling on Tuesday is more likely to engage in a risky and dangerous situation than other days in a week. Specifically, the probability of being caught in the fatal situation on Tuesday is at least twice higher than any other days.
2) Most cyclist crashes occurred during daytime, and the model results found that in Liverpool, riding at daytime or night with street lights, has a similar probability with dark without light, which are different with the findings from previous studies in the field.

5.4 Practical contributions

In this section, we emphasise the findings which help explain how the proposed model and research implications can aid risk informed safety policy making and control measure development.

5.4.1 Policy making

It is the responsibility of transport authorities to develop feasible safety policies to ensure cycling safety at national and regional levels. For example, the Department for Transport proposed an instruction (STATS20) in 2011 for police forces and local transport authorities around UK to guide them regulating the behaviour of road users, including cyclists. The findings from our research guide the Liverpool city region to develop its cycling safety measures. When formulating relevant policies, 1st priority variables (age, district, day, encountering vehicle type, first point of impact) and some 2nd priority variables (according to city conditions and costs) are taken into account in their policy making process. Education lessons become compulsory for personnel engaging in dangerous cycling and driving and more separate lanes and traffic lights near junctions are built in the Liverpool city region.

Furthermore, the transport systems in different cities vary, indicating the cycling safety policymaking in different cities should be developed with respect to the unique features of their local conditions. The methodology can be used to develop BN risk models for different cities to find distinct features for city specific safety policies. For instance, according to the analysis result in Section 5.3, motorcyclist behaviour should be better regulated and their safety education should be enhanced in the Liverpool city region.

5.5 Limitations and further improvements

Although the research provides some valuable insights to improve cycling safety, limitations still exist. Further studies are needed to improve the research from the following aspects:

1) The proposed model is not universal– The accident data collected and trained in this research is from the Liverpool region. It means the obtained outputs this study need to be verified before their direct applications in other regions. However, the research methodology proposed in this

research is generic and can be used widely In fact, more applications of the proposed methodology could help generalise the findings that currently fits the Liverpool region only.

2) The state definition of risk variables in the proposed model could be adjusted to be more specific when a large scale accident dataset becomes available. For instance,

- the location will be specifically expressed by roads or postcodes.
- The states of ‘time’ can be changed from the current rush/non-rush hours into hourly grades.

Obviously, such investigation needs the support of big data. It involves data mining, training, and high-level computation, requiring a large number of accident reports as raw data.

3) Some observations and findings (i.e. those discussed in Section 5.3.2 such as high risk on Tuesday), need to be further investigation.

6. Conclusions

To address the possible emerging risks introduced by increasing cycling traffic in cities, a new BN cycling risk model is developed to analyse severity of cycling accidents in this paper. Based on 5-year full road accident reports by Merseyside Police, the risk factors influencing the accident severity are first identified and classified. Through TAN learning, the data-driven BN model is developed to aid the analysis and prediction of cycling accident severity. The model and results are validated using real accident data from 2018 in the Liverpool city region and sensitivity analysis. The proposed methodology can be generalized for its applications in other regions when and where the accident data are available to improve cycling safety in a large scope.

The findings include that the risk factors are categorised into three types: 1st priority variables (i.e. cyclist age, district, day, encountering vessel type), 2nd priority variables (i.e. road type, V4, O4), and low priority variables (i.e. street lighting, road surface). Some specific features of the severity of cycling accidents in the Liverpool region are relating to high risk when accidents occur with motorcycles and on Tuesday and low risk when they happen at junctions, with dry road surface and involve female cyclists.

Such findings help transport authorities to rationalise their safety policy making in cycling practice. In the case study in this paper, the results have been used for policy making in Liverpool. For instance, education lessons are required for personnel engaging in dangerous cycling and driving and separate cycling lanes and traffic lights near junctions are built in the Liverpool city region.

Further effort will be needed to improve the conceptual risk prediction model by considering more specific state definition of risk variables, the combined impact of multiple variables/hazards (three or more), and the development of cost-effective measures to control cycling risks.

Acknowledgement

This research was financially supported by Mersey Travel and Department for Transport UK through two linked projects (i.e. Cycling Hazard Identification in the Liverpool City Region and Advanced Model for Cycling Risk Analysis and Prediction). We also thank Merseyside Police for providing us with research primary data.

References

Altman, D. G., 1999. Practical statistics for medical research. New York, NY: Chapman & Hall/CRC Press.

Andersson, A-L. & Bunketorp, O. 2002. Cycling and alcohol. Injury International Journal of the Care of the Injured, Vol. 33, pp. 467-471

Anstey, K.J., Wood, J., Lord, S., Walker, J.G. 2005. Cognitive, sensory and physical factors enabling driving safety in older adults. Clinical Psychology Review, Vol. 25, pp. 45–65

Bacchieri, G., Barros, A., dos Santos, J.V. & Gigante, D.P. 2010. Cycling to work in Brazil: Users profile, risk behaviors, and traffic accident occurrence. Accident Analysis and Prevention, Vol. 42, pp. 1025-1030

Behnood, A., Mannering, F. 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research, Vol. 16, pp. 35-47

Bíl, M., Bílováa, M., Müllerb, I. 2010. Critical factors in fatal collisions of adult cyclists with automobiles. Accident Analysis and Prevention, Vol. 42, pp. 1632–1636

Boufous, S., de Rome, L., Senserrick, T. & Ivers, R.Q. 2017. Single-versus multi-vehicle bicycle road crashes in Victoria, Australia. Injury Prevention, Vol. 19, Issue 5, pp. 358-363

Buntine, W. 1991. Theory refinement in bayesian networks. Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence, pp. 52-60

Chen, P. 2016. Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. Safety Science, Vol. 79, pp. 336-343.

Cooper, G. F. & Herskovits, E. 1992. A Bayesian method for the induction of probabilistic networks from data. Machine Learning, Vol 9(4), pp. 309-347.

Dardari, D., Decarli, N., Guerra,A., Al-Rimawi,A., Puchades, V.M., Prati, G., De Angelis, M., Fraboni, F. & Pietrantoni, L. 2017. High-Accuracy Tracking Using Ultra Wideband Signals for Enhanced Safety of Cyclists. Mobile Information Systems, 2017, pp. 1-13

Davison, K.K. & Lawson, C.T. 2006. Do attributes in the physical environment influence children's physical activity? A review of the literature. International Journal of Behavioral Nutrition and Physical Activity, pp 3-19

Dubbeldam, R., Baten, C., Buurke, J.H. & Rietman, J.S. 2017. SOFIE, a bicycle that supports older cyclists? Accident Analysis and Prevention, Vol. 105, pp. 107-123

Friedman, N., Geiger, D., Goldszmitd, M., 1997, Bayesian network classifiers. Machine learning, Vol. 29, 131-163

Ghekiere, A., Van Cauwenberg, J., de Geus, B., Clarys, P., Cardon, G. Salmon, J., De Bourdeaudhuij, I. & Deforche, B. 2014. Critical Environmental Factors for Transportation Cycling in Children: A Qualitative Study Using Bike-Along Interviews. PLoS ONE, Vol. 9, Issue 9, pp. 1-10

Hänninen, M. & Kujala, P. 2012. Influences of variables on ship collision probability in a Bayesian belief network model. Reliability Engineering & System Safety, Vol. 102, pp. 27-40

Higgins, P. 2005. Exercise-based transportation reduces oil dependence, carbon emissions and obesity. Environmental Conversation, Vol. 32, Issue 3, pp. 197-202

Heinen, E., van Wee, B. & Maat, K. 2010. Commuting by Bicycle: An Overview of the Literature. Transport Reviews, Vol. 30, Issue 1, pp. 59-96

Heydari, S., Fu, L., Mirandan-Moreno, L.F., Jopseph, L. 2017. Using a flexible multivariate latent class approach to model correlated outcomes: A joint analysis of pedestrian and cyclist injuries. Analytic Methods in Accident Research, pp. 16-27.

Jain, Anil K., Dubes, Richard C. 1988. Algorithms for Clustering Data. Prentice Hall.

Juhra, C., Wieskotter, B., Chu, K., Trost, L., Weiss, U., Messerschmidt, M., Malczyk, A., Heckwolf, M. & Raschke, M. 2012. Bicycle accidents – Do we only see the tip of the iceberg? A prospective multi-centre study in a large German city combining medical and police data. Injury, International Journal of Care Injured, Vol. 43, pp. 2026-2034

Krause, J., Small, M.J., Haas, A. & Jaeger, C.C. 2016. An expert-based Bayesian assessment of 2030 German new vehicle CO2emissions and related costs. Transport Policy, Vol. 52, pp. 197-208

Kondo M.C., Morrison, C., Guerra, E., Kaufman, E.J. & Wiebe, D.J. 2018. Where do bike lanes work best? A Bayesian spatial model of bicycle lanes and bicycle crashes. Safety Science, Vol. 103, pp. 225-233.

Langley, P., Iba, W. & Thompson, K. 1992. An analysis of Bayesian classifiers. San Jose, CA, USA, s.n., pp. 223-238.

Li, K. X., Yin, J., Bang, H. S., Yang, Z., Wang, J., 2014, Bayesian network with quantitative input for maritime risk analysis. Transportmetrica A: Transport Science, Vol. 10, Issue 2, pp. 89-118

Manton, R., Rau, H., Fahy, F. & Sheahan, J. 2016. Using mental mapping to unpack perceived cycling risk. Accident Analysis and Prevention, Vol. 88, pp. 138-149

Merriam-Webster Online Dictionary, 2008. Cluster analysis. http://www.merriam-webster-online.com

Murphy, P. M. & Aha, D. W. 1995. UCI repository of machine learning database. [Online] Available at: http:// www.ics.uci.edu/mlearn/MLRepository.html

Oke O., Bhalla K., Love D.C. & Siddiqui S. 2015. Tracking global bicycle ownership patterns. Journal of Transport and Health, Vol. 2, Issue 4, pp. 490-501.

Olkkonen, S. & Honkanen, R. 1990. The role of alcohol in nonfatal bicycle injuries. Accident Analysis and Prevention, Vol. 22, Issue 1, pp. 89-96

Osama, A., Sayed, T. 2017. Investigating the effect of spatial and mode correlations on active transportation safety modeling. Analytic Methods in Accident Research, 16, pp. 60-74.

Oteniya, L. 2008. Bayesian Belief Networks for Dementia Diagnosis and Other Applications: A Comparison of Hand-Crafting and Construction using A Novel Data Driven Technique. Department of Computing Science and Mathematics, University of Stirling, Scotland, Technical Report CSM-179, ISSN 1460-9673

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A., Jarawan, E., Mathers, C. 2014. World report on road traffic injury prevention. World Health Organization, Geneva

Pucher, J., Dill, J. & Handy, S. 2010. Infrastructure, programs, and policies to increase bicycling: An international review. Preventive Medicine, Vol. 50, pp. 106-125

Pucher, J. & Dijkstra, L. 2003. Promoting Safe Walking and Cycling to Improve Public Health: Lessons from the Netherlands and Germany. American Journal of Public Health, Vol. 93, Issue 9, pp. 1509-1516

Puchades, V.M., Fassina, F., Fraboni, F., De Angelis, M., Prati, G., de Waard, D. & Pietrantoni, L. 2018. The role of perceived competence and risk perception in cycling near misses. Safety Science, Vol. 105, pp. 167-177

Quinlan, J. R., 1995. C4.5: Programs for machine learning. San Francisco, CA: Morgan Kaufmann.

Rämä, P. 2001. Effect of weather-controlled variable message signing on driver behaviour. Technical Research Centre of Finland.

Reynolds, C., Harris, M.A., Teschke, K., Cripon, P.A. & Winters, M. 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environmental Health, Vol. 8, Issue 47, pp. 1-19

Robert, L. Thorndike. 1953. Who belongs in the family? Psychometrika, Vol. 18, pp. 267–276

Rodgers, G.B. 1995. Bicyclist deaths and fatality risk patterns. Accident Analysis & Prevention, Vol. 27, Issue 2, pp. 215-223

Rodgers, G.B. 1997. Factors Associated with the Crash Risk of Adult Bicyclists. Journal of Safety Research, Vol. 28, Issue 4

Serrano, B.M., Gonzalez-Cancelas, N., Soler-Flores, F. & Camarero-Orive, A. Classification and prediction of port variables using Bayesian Networks. Transport Policy, Vol. 67, pp. 57-66

Schepers, P., Hagenzieker, M., Methorst, R., van Wee, B. & Wegman, F. 2014. A conceptual framework for road safety and mobility applied to cycling safety. Accident Analysis & Prevention, Vol. 62, pp. 331-340

Tin, S.T., Woodward, A., Thornley, S., Langley, J., Rodgers, A. & Ameratunga, S. 2009. Cyclists' attitudes toward policies encouraging bicycle travel: findings from the Taupo Bicycle Study in New Zealand. Health Promotion International, Vol. 25, Issue 1, pp. 54-62

UK DfT. 2020. STATS19 road accident injury statistics – report form

UK DfT, 2020, Instructions for the Completion of Road Accident Reports from non-CRASH Sources, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/230596/stats20-2011.pdf. Accessed on 11 December 2020.

Vandenbulcke, G., Thomas, I. & Panis, L.I. 2014. Predicting cycling accident risk in Brussels: A spatial case-control approach. Accident Analysis and Prevention, Vol. 62, pp. 341-357

Wang, L.K., Yang, Z.L. 2018. Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. Reliability Engineering and System Safety, Vol. 180, pp. 277-289

Wood, J.M., Lacherez, P.F., Marszalek, R.P. & King, M.J. 2009. Drivers and cyclists' experiences of sharing the road: Incidents, attitudes and perceptions of visibility. Accident Analysis and Prevention, Vol. 41, pp. 772-776

Xie, Y., Lord, D. & Zhang, Y. 2007. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accident Analysis & Prevention, Vol. 39, Issue 5, pp. 922-933

Yang, M., Wu, J., Rasouli, S., Cirillo, C. & Li, D. 2017. Exploring the impact of residential relocation on modal shift in commute trips: Evidence from a quasi-longitudinal analysis. Transport Policy, Vol. 59, pp. 142-152

Yang, Z.S, Yang, Z.L. & Yin, J. 2018. Realizing advanced risk-based Port State Control Inspection using data-driven Bayesian networks. Transportation Research Part A: Policy and Practice, Vol. 110, pp. 38-56

Yang, Z.S., Yang, Z.L., Yin, J. & Qu, Z. 2018. A risk-based game model for rational inspections in Port State Control. Transportation Research Part E: Logistics and Transportation Review, Vol. 118, pp. 477-495

Yang Z., Bonsall, S., Wang, J., 2009, Use of fuzzy evidential reasoning in maritime security assessment, Risk Analysis, Vol. 29, Issue 1, pp. 95-120

Zhang, D., Yan, X. P., Yang, Z. L., Wall, A., & Wang, J. 2013. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. Reliability Engineering & System Safety, Vol. 118, pp. 93–105

Appendix 1 Hazards in cycling safety in literature

1) Cyclist behaviour and personal characteristics

Many bicycle-related crashes and accidents nowadays are caused by inappropriate behaviours of cyclists. As human factor is one of the main causes of road accidents, cyclist behaviours also play an important role in bicycle crashes (Schepers et al., 2014; Behnood & Mannering, 2017). Figure A1.1 lists the identified hazards in this category with respect to their appearance frequencies in previous studies.

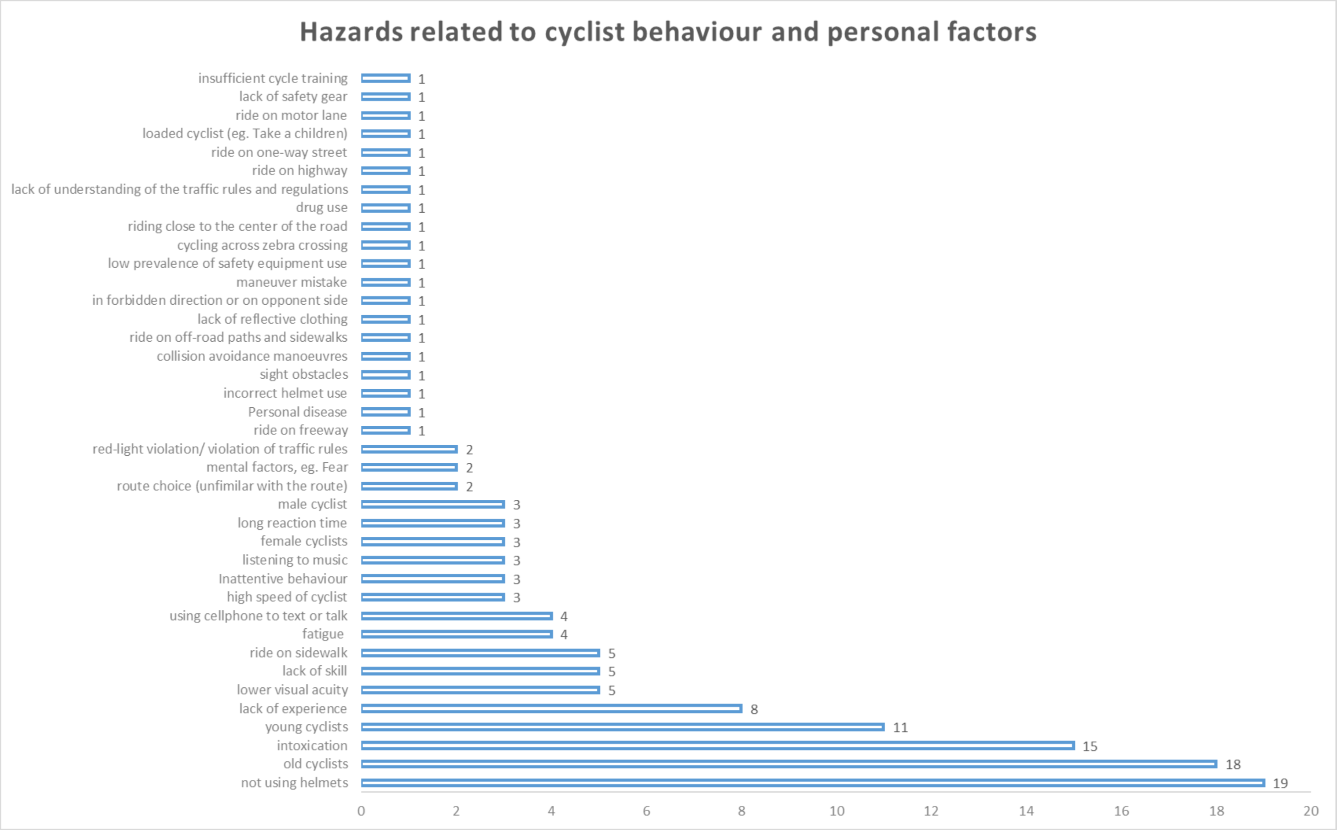


Figure A1.1. Hazards in cyclist behaviour and personal characteristics

2) Environmental conditions

The environment is known to be one of the influencing aspects to cycling safety (Davison & Lawson, 2006). The hazards generated by environmental conditions are critical factors (see Figure A1.2) influencing cycling safety (Ghekiere et al., 2014).

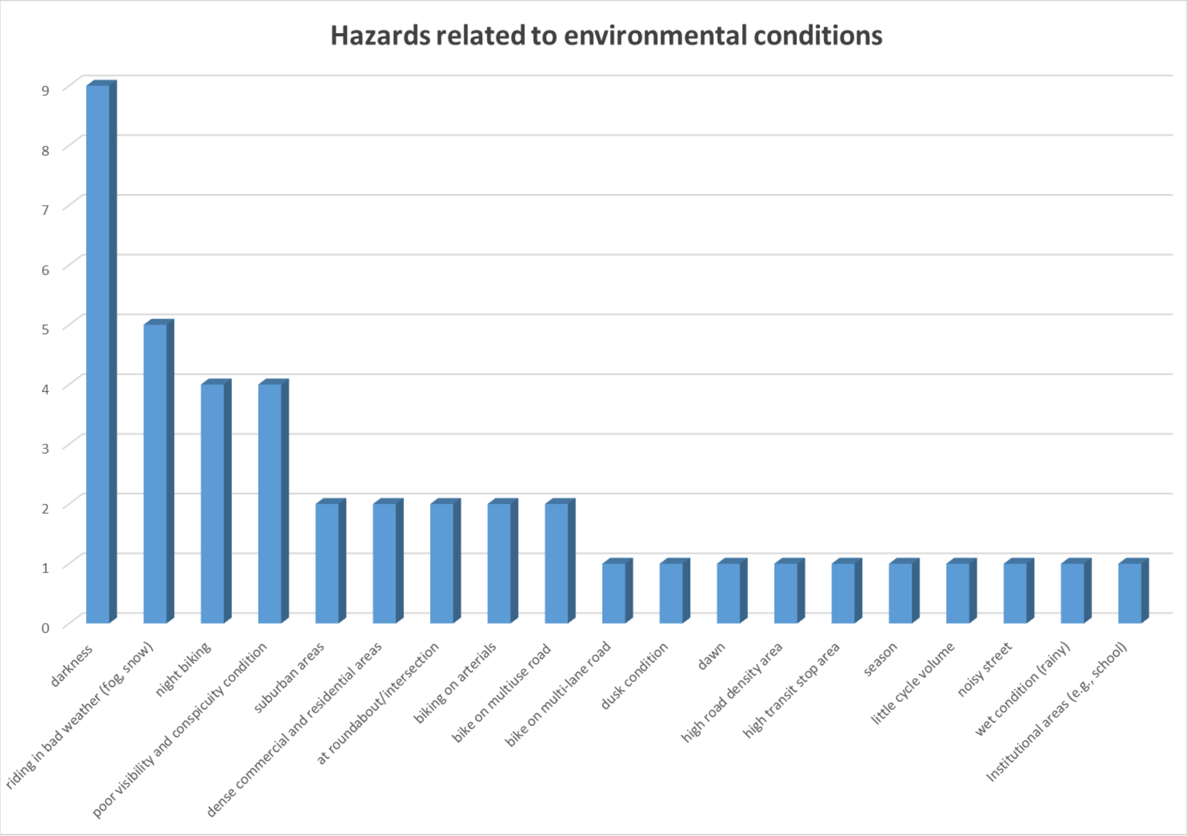


Figure A1.2 Hazards in ‘Environmental conditions’

3) Road infrastructure issue

Transport for London (2008) pointed out that cyclists tend to report fear of injury from lack of specialized cycling infrastructure, e.g. the segregated cycling routes. From the identified hazards in this category in Figure A1.3, it is not surprising to see that one of the key approaches to reducing the fear and risk of injury for cyclists is through engineering means and, in particular, through designated cycling transport infrastructure (Rodgers, 1997).

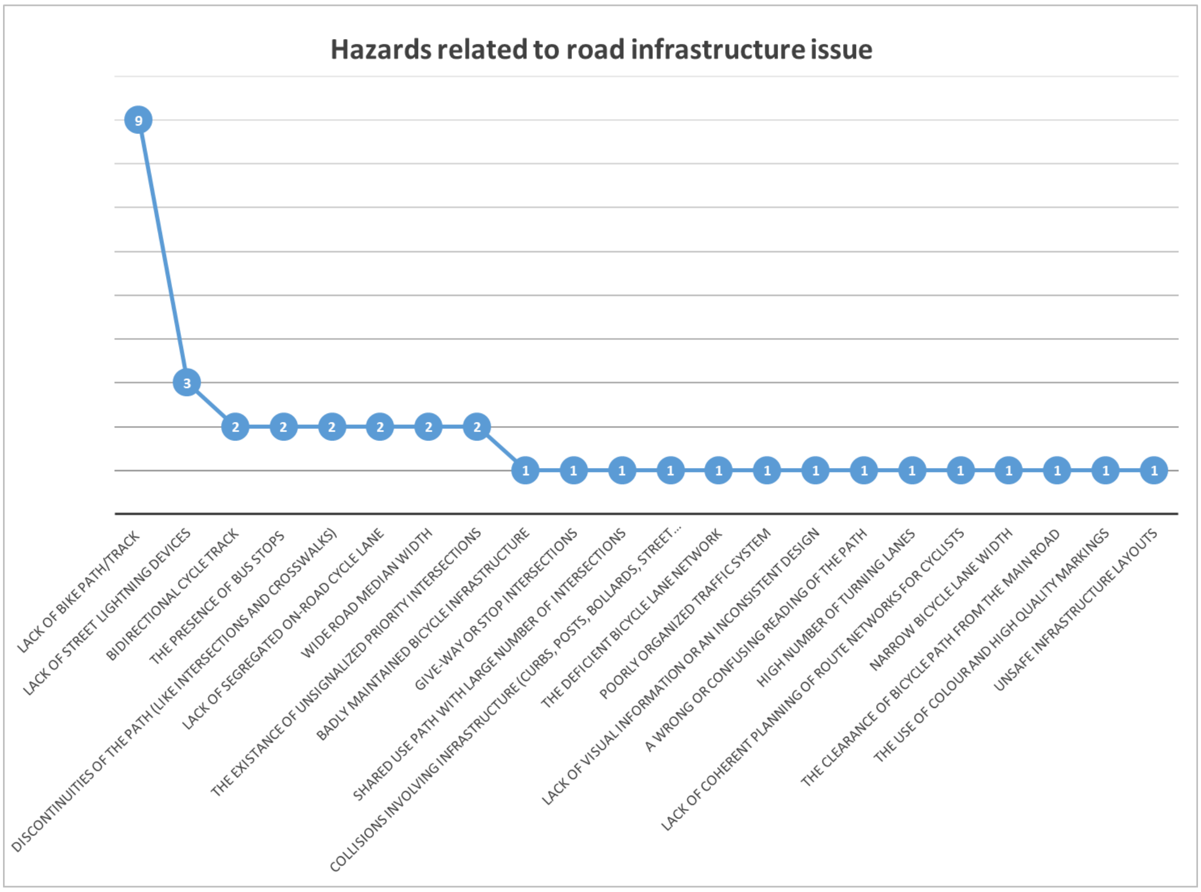


Figure A1.3. Hazards in ‘Road Infrastructure Issue’

4) Interaction with other road users

The hazard of ‘collisions with vehicles’ is a major concern, which has been evidenced in the findings from many studies (Olkkonen et al., 1990; Rodgers, 1995). According to a conceptual framework proposed for road and mobility safety (Schepers et al., 2014), interactions and collisions with other road users are risks resulting from travel characteristics (see Figure A1.4).

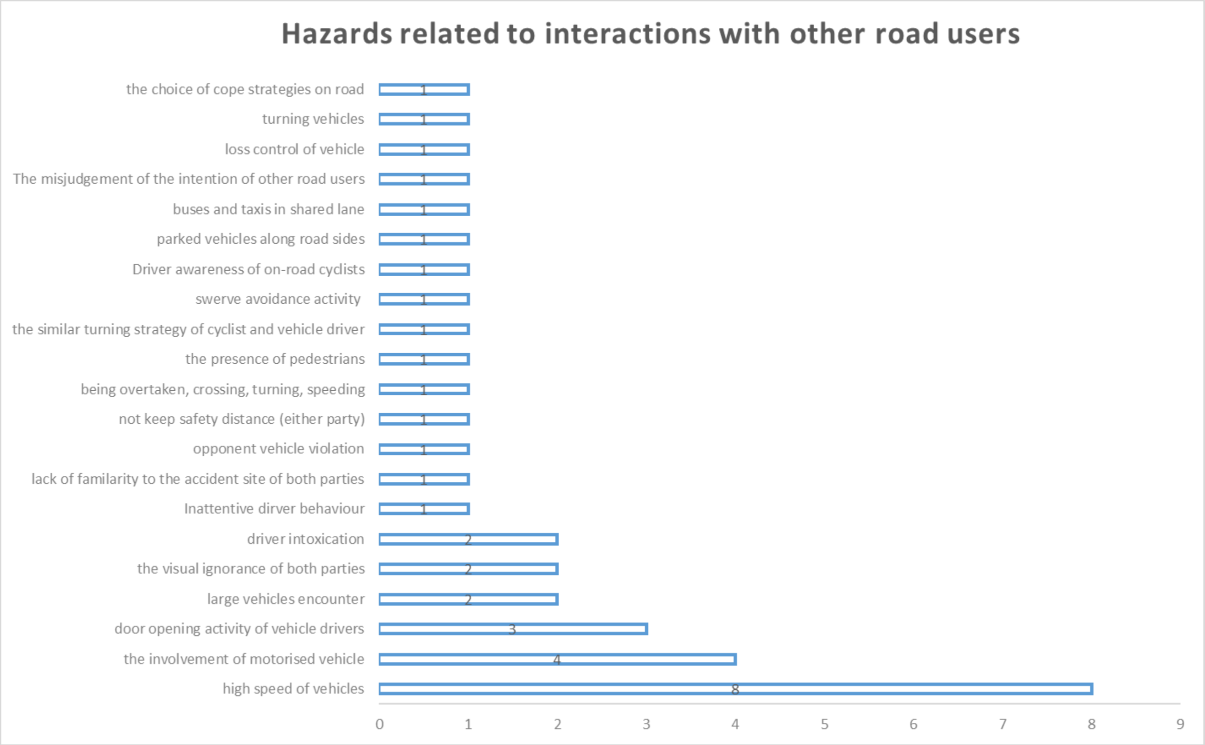


Figure A1.4. Hazards in ‘Interaction with other road users’

5) Hazardous road conditions

Generally, hazardous road conditions refer to the physical and traffic conditions of a road that can cause harm to cyclists. Therefore, the hazards in this category are analysed from two aspects: the physical hazardous road condition (e.g. road surface condition) and road traffic management (e.g. traffic congestion) in Figure A1.5.

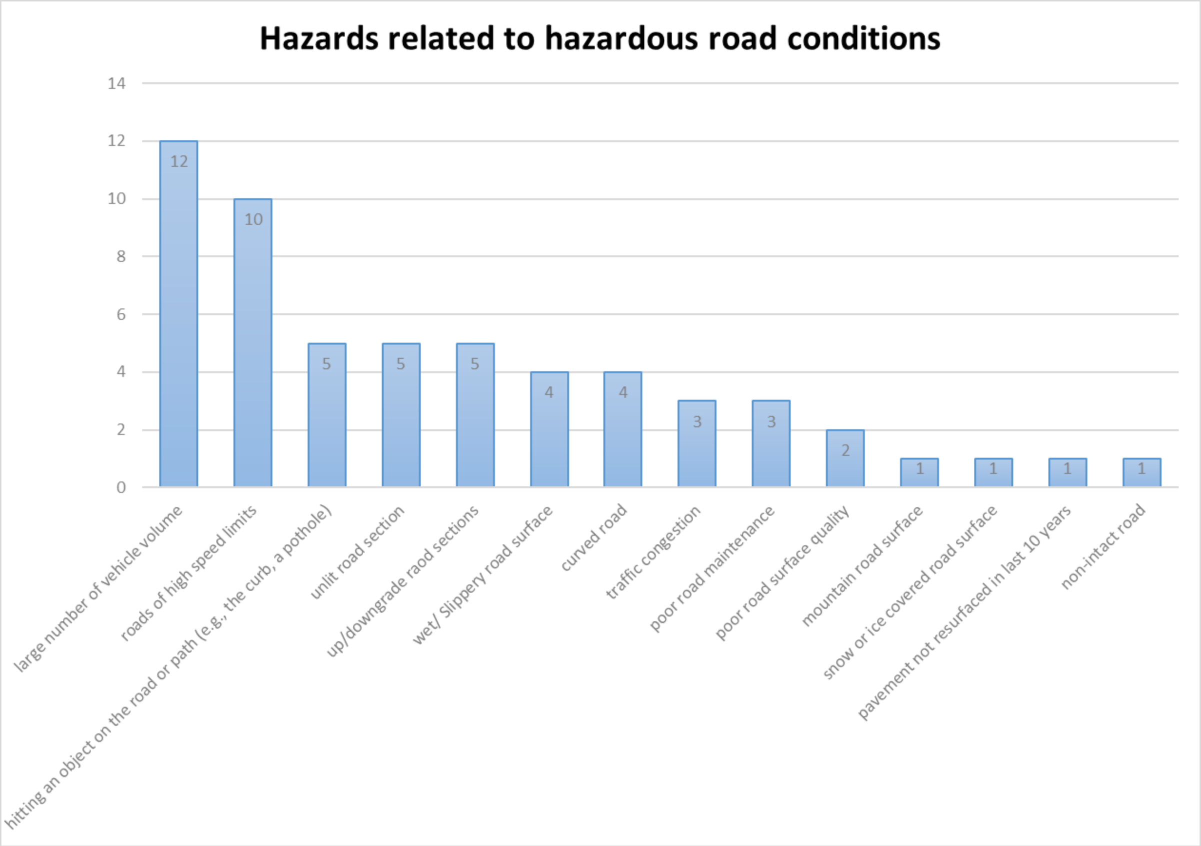


Figure A1.5. Hazards in ‘Hazardous Road Condition’

6) Bike-related factors

Failures of a bicycle also affect the safety of cyclists. For example, in Pelotas, a southern city in Brazil, almost 30% of bicycles lack of effective brakes, causing high-frequent cycling incidents/accidents (Bacchieri et al., 2010). Meanwhile, poor conditioned bicycles such as poorly lubricated, lacked gears, poorly maintained, and unlikely to reach high speeds cause safety concerns in cycling (e.g. Wood et al., 2009; Bacchieri et al., 2010; Tin et al., 2009; Dubbeldam et al., 2017).

Appendix 2 Variable identification and screening process

1) Selected contributory factors for model construction

The classification of the contributory factors is conducted to eliminate the trivial factors from the BN model by the following filtering rules. As a result, the contributory factors are classified into 16 categories, as shown in Table A2.1.

- The factors that appear in over 10% of the accident reports are remain individually.
- Other remained factors are classified and merged into different categories according to the classification defined in the STATS20 handbook when causing cycling accidents.

Table A2.1 Classification of contributory factors

	Appearance (V)	Slight C	Serious C	Fatal C	Appearance (O)	Slight C	Serious C	Fatal C
1.Road Environment issue (101-110)	23	15	8	0	35	28	7	0
2.Non-motor vehicle defects (201-206)	45	32	13	0	9	7	2	0
3.Disobeying the traffic facilities & rules (301-304)	49	31	18	0	80	57	22	1
4.Injudicious action of cyclist/driver on road (305-310)	264	185	75	4	85	54	28	3
5.Failed to judge other person's path/speed (406)	131	88	42	1	310	243	66	1
6.Other Judgment failures/errors (401, 402, 404, 407)	41	30	11	0	230	176	53	1
7.Poor turn maneuver (403)	60	41	18	1	224	170	52	2
8.Failed to look properly (405)	412	303	109	0	974	744	227	3
9.Reaction to unexpected cases (408-410)	52	31	21	0	31	20	10	1
10.Physical/mental illness or impairment (501-505)	38	24	13	1	18	10	8	0
11.Distraction/Misleading event (506-510)	106	71	35	0	33	28	5	0
12.Emotional driving/riding (601-603)	89	63	25	1	170	133	36	1
13.Inexperience of driver/rider (605-607)	15	10	5	0	13	9	4	0
14.Obscured vision (701-710)	41	33	8	0	163	132	30	1
15.Pedestrian related factors (801-810)	26	18	8	0	29	24	5	0
16.Other factors (901-904, 999)	8	5	3	0	71	54	15	2

**The numbers behind each category are the code number of original contributory factors in STATS19. For example, 'Road environment issue' consists of original factors from 101 to 110.*
*** 'V' represents victim, 'O' represents other road users, 'C' represents the consequence*

2) Removal of insignificant other factors

Besides contributory factors, the STATS19 accident reports contain other factors, some of which are less relevant to this study (given the report is designed to record all types of road accidents with no specific reference to cycling). These factors are viewed as insignificant variables. In this section, the insignificant variables in STATS19 reports are shown as follows and the reasons for excluding them as the nodes in the BN are presented accordingly when it is necessary.

- Police officer attendance

It means whether a police officer attended the scene of the accident and obtained the details for this report. It has little relevance and impact to the accident severity as a typical post-accident activity.

- Occurrence date

According to the statistical analysis, there is insignificant difference in terms of the frequency of different types of accident severity between different months and years. Furthermore, the impact of the occurrence date on accident severity is generated indirectly through other date related factors such as weather and daytime/darkness. Since these ‘date-related’ factors have been included in BN construction separately, ‘occurrence date’ is excluded from the model.

- Other factors that have been addressed in full or part by the included contributory factors include Pedestrian crossing with Human control & Physical facilities, Special site conditions, Object in carriageway, Hit objective, Breath test.

Pedestrian crossing with Human control & Physical facilities – Contributory factor 15 ‘Pedestrian related factors’: this factor refers to whether there exist crossing facilities at the location of accident, i.e. zebra crossing, footbridge or subway, central refuge.

Object in carriageway & Hit Objective – Contributory factor 1 ‘Road environmental issue': this factor refers to those objects that are not expected to be found on the road

Special conditions at site– Contributory factor 1 ‘Road environmental issue': this factor refers to whether there were special conditions at the accident site, i.e., Automatic traffic signal out, roadworks, mud.

Breath test – Contributory factor 10 ‘Physical/mental illness or impairment’: this factor refers to the test used to check whether the driver is drunk

Junction location at impact – Junction details of accident: compared with ‘Junction location at impact’, ‘Junction details of accident’ is a more comprehensive factor. It not only describes the location of an accident, but also clarify the junction type of which the accident happens.

- Journey purpose

According to the collected accident reports, the journey purposes of over 70% cycling accidents remain unclear, hence it has not been set as a node in the BN. Furthermore, from our literature search, it indicates that there is no strong correlation between journey purpose and cycling accident severity.

Appendix 3 Comparison between the real results and theoretical results (based on Year 2018)

No	Severity	Model delivery	P(slight)	P(Serious)	P(Fatal)	No	Severity	Model delivery	P(slight)	P(Serious)	P(Fatal)
1800153	Fatal	Fatal	5.2	1.7	93.1	1801204	Slight	Slight	73.2	26.8	0
1801099	Serious	Serious	15.5	84.5	0	1801205	Slight	Slight	76.6	23.4	0
1800215	Serious	Serious	20.2	79.8	0	1801223	Slight	Slight	83.9	16.1	0
1800277	Serious	Serious	36.5	63.5	0	1801238	Slight	Slight	75.8	24.2	0
1800547	Serious	Serious	43.4	56.6	0	1801241	Slight	Slight	98.6	1.4	0
1800568	Serious	Serious	19	81	0	1801260	Slight	Slight	76.6	23.4	0
1800750	Serious	Serious	46.3	53.7	0	1801261	Slight	Slight	100	0	0
1800866	Serious	Serious	39	61	0	1801279	Slight	Slight	69.2	30.8	0
1800894	Serious	Serious	19.1	80.9	0	1801283	Slight	Slight	89.7	10.3	0
1801130	Serious	Serious	36.1	63.9	0	1801295	Slight	Slight	87.1	12.9	0
1801176	Serious	Serious	45.8	54.2	0	1801313	Slight	Slight	71.5	28.5	0
1801203	Serious	Serious	46.6	53.4	0	1801325	Slight	Slight	84.9	15.1	0
1801212	Serious	Serious	33.6	66.4	0	1801327	Slight	Slight	95.6	4.4	0
1801342	Serious	Serious	46.5	53.5	0	1801328	Slight	Slight	57.9	42.1	0
1801436	Serious	Serious	0.1	99.9	0	1801334	Slight	Slight	99.9	0.1	0
1801479	Serious	Serious	24	76	0	1801370	Slight	Slight	83.7	16.3	0
1801511	Serious	Serious	20.3	79.7	0	1801377	Slight	Slight	88.5	11.5	0
1801587	Serious	Serious	42.6	57.4	0	1801380	Slight	Slight	99.5	0.5	0
1801699	Serious	Serious	46	54	0	1801393	Slight	Slight	92	8	0
1801778	Serious	Serious	0.9	99.1	0	1801396	Slight	Slight	55.8	44.2	0
1801985	Serious	Serious	35.1	64.9	0	1801413	Slight	Slight	90.2	9.8	0
1801896	Serious	Serious	37.6	62.4	0	1801420	Slight	slight	55.6	44	0.4
1802001	Serious	Serious	47	53	0	1801440	Slight	slight	86.8	13.2	0
1802055	Serious	Serious	43.2	56.8	0	1801457	Slight	slight	90.4	9.6	0
1802103	Serious	Serious	1.2	98.8	0	1801461	Slight	slight	89.6	10.4	0
1802220	Serious	Serious	37.6	62.4	0	1801466	Slight	slight	82.6	17.4	0
1802244	Serious	Serious	46.8	53.2	0	1801473	Slight	slight	76.1	23.9	0
1802290	Serious	Serious	49.7	50.3	0	1801483	Slight	slight	80.8	19.2	0
1802351	Serious	Serious	44.4	55.6	0	1801497	Slight	slight	69	31	0
1800478	Serious	Serious	31.4	68.6	0	1801499	Slight	slight	60.8	39.2	0
1800980	Serious	Serious	44.1	55.9	0	1801504	Slight	slight	91.4	8.6	0
1801102	Serious	Serious	45.6	54.4	0	1801510	Slight	slight	68.1	31.9	0
1801256	Serious	Serious	17.6	82.4	0	1801514	Slight	slight	76.3	23.7	0
1800936	Serious	Slight	99.7	0.3	0	1801527	Slight	slight	99.3	0.7	0
1801001	Serious	Slight	75.7	24.3	0	1801530	Slight	slight	56.5	43.5	0
1801759	Serious	Slight	82.8	17.2	0	1801534	Slight	slight	89.8	10.2	0
1802374	Serious	Slight	91.6	8.4	0	1801544	Slight	slight	66.9	33.1	0
1800549	Slight	Serious	32	68	0	1801552	Slight	slight	61.9	36.9	1.2
1800574	Slight	Serious	39.4	60.6	0	1801555	Slight	slight	83.9	16.1	0
1800643	Slight	Serious	38.2	61.8	0	1801564	Slight	slight	79.2	20.8	0

1801018	Slight	Serious	43.4	56.6	0	1801576	Slight	slight	86.4	13.6	0
1801047	Slight	Serious	33	67	0	1801584	Slight	slight	96	4	0
1801418	Slight	Serious	45.7	54.3	0	1801590	Slight	slight	84.6	15.4	0
1801472	Slight	Serious	34.7	65.3	0	1801596	Slight	slight	74.3	25.7	0
1801875	Slight	Serious	46.4	53.6	0	1801618	Slight	slight	89.7	10.3	0
1801900	Slight	Serious	11.4	88.6	0	1801621	Slight	slight	83.6	16.4	0
1802296	Slight	Serious	33.5	66.5	0	1801625	Slight	slight	83.7	16.3	0
1802335	Slight	Serious	1.6	98.4	0	1801629	Slight	slight	70.4	29.6	0
1800029	Slight	Slight	99.7	0.28	0.02	1801675	Slight	Slight	83.2	16.8	0
1800079	Slight	Slight	93.2	6.85	-0.05	1801710	Slight	Slight	78.7	21.3	0
1800113	Slight	Slight	88.2	11.8	0	1801713	Slight	Slight	67.2	32.8	0
1800130	Slight	Slight	100	0	0	1801714	Slight	Slight	69.9	30.1	0
1800176	Slight	Slight	89.6	10.4	0	1801723	Slight	Slight	96.2	3.8	0
1800179	Slight	Slight	72.2	27.8	0	1801731	Slight	Slight	71	29	0
1800185	Slight	Slight	65	35	0	1801732	Slight	Slight	93.7	3.8	0
1800186	Slight	Slight	65.1	34.9	0	1801783	Slight	Slight	78.8	21.2	0
1800187	Slight	Slight	95.4	4.6	0	1801794	Slight	Slight	87.1	12.9	0
1800191	Slight	Slight	90.97	9.03	0	1801800	Slight	Slight	90.7	9.3	0
1800197	Slight	Slight	100	0	0	1801804	Slight	Slight	68.5	31.5	0
1800218	Slight	Slight	63.3	36.7	0	1801809	Slight	Slight	86.7	13.3	0
1800219	Slight	Slight	88.7	11.3	0	1801872	Slight	Slight	100	0	0
1800221	Slight	Slight	76.6	23.4	0	1801876	Slight	Slight	97.4	2.6	0
1800252	Slight	Slight	88.4	11.6	0	1801887	Slight	Slight	83.5	16.5	0
1800263	Slight	Slight	81.9	18.1	0	1801888	Slight	Slight	69.6	30.4	0
1800327	Slight	Slight	65.9	34.1	0	1801890	Slight	Slight	86.4	13.6	0
1800334	Slight	Slight	93.5	6.5	0	1801905	Slight	Slight	84	16	0
1800375	Slight	Slight	93.9	6.1	0	1801911	Slight	Slight	75.3	24.7	0
1800386	Slight	Slight	92.1	7.9	0	1801912	Slight	Slight	100	0	0
1800401	Slight	Slight	68	32	0	1801923	Slight	Slight	100	0	0
1800487	Slight	Slight	74.2	25.8	0	1801925	Slight	Slight	79.2	20.8	0
1800509	Slight	Slight	91.8	8.2	0	1801943	Slight	Slight	100	0	0
1800529	Slight	Slight	88.2	11.8	0	1801956	Slight	Slight	55.1	44.9	0
1800536	Slight	Slight	99.9	0.1	0	1801974	Slight	Slight	75.4	24.6	0
1800544	Slight	Slight	100	0	0	1801997	Slight	Slight	88.6	11.4	0
1800545	Slight	Slight	100	0	0	1802006	Slight	Slight	69.7	30.3	0
1800594	Slight	Slight	66.9	33.1	0	1802008	Slight	Slight	55.2	44.8	0
1800623	Slight	Slight	82.9	17.1	0	1802015	Slight	Slight	52.5	47.5	0
1800627	Slight	Slight	92.5	7.5	0	1802020	Slight	Slight	91.7	8.3	0
1800630	Slight	Slight	88.1	11.9	0	1802034	Slight	Slight	54.3	45.7	0
1800642	Slight	Slight	63.4	36.6	0	1802043	Slight	Slight	86.9	13.1	0
1800679	Slight	slight	99.6	0.4	0	1802059	Slight	Slight	79.5	20.5	0
1800696	Slight	slight	99.8	0.2	0	1802069	Slight	Slight	99.2	0.8	0
1800719	Slight	slight	71.7	28.3	0	1802113	Slight	Slight	78	22	0
1800745	Slight	slight	51.5	48.5	0	1802128	Slight	Slight	100	0	0

1800769	Slight	slight	87.8	12.2	0	1802145	Slight	Slight	80.2	19.8	0
1800850	Slight	slight	83.3	16.7	0	1802154	Slight	Slight	64.4	35.6	0
1800855	Slight	slight	97.5	2.5	0	1802156	Slight	Slight	99.8	0.2	0
1800874	Slight	slight	76.9	23.1	0	1802160	Slight	Slight	88.1	11.9	0
1800895	Slight	slight	63.2	36.8	0	1802180	Slight	Slight	60.1	39.9	0
1800943	Slight	slight	98.6	1.4	0	1802204	Slight	Slight	78.2	21.8	0
1800944	Slight	slight	41.1	58.9	0	1802207	Slight	Slight	99.3	0.7	0
1800947	Slight	slight	93.7	5.85	0.42	1802208	Slight	Slight	93.7	6.3	0
1800953	Slight	slight	100	0	0	1802213	Slight	Slight	64.6	35.4	0
1800961	Slight	slight	88.2	11.8	0	1802216	Slight	Slight	78.9	21.1	0
1800973	Slight	slight	99.4	0.6	0	1802276	Slight	Slight	100	0	0
1800985	Slight	slight	70.5	29.5	0	1802281	Slight	Slight	83.2	16.8	0
1801023	Slight	Slight	83.5	16.5	0	1802302	Slight	Slight	78.1	21.9	0
1801025	Slight	Slight	82.2	17.8	0	1802307	Slight	Slight	100	0	0
1801029	Slight	Slight	81.8	18.2	0	1802314	Slight	Slight	75.2	24.8	0
1801062	Slight	Slight	76.3	23.7	0	1802320	Slight	Slight	96.2	3.8	0
1801104	Slight	Slight	94.2	5.8	0	1802325	Slight	Slight	69.9	30.1	0
1801106	Slight	Slight	83.9	16.1	0	1802329	Slight	Slight	99.9	0.1	0
1801139	Slight	Slight	75.3	24.7	0	1802337	Slight	Slight	89.9	10.1	0
1801152	Slight	Slight	53.4	46.6	0	1802349	Slight	Slight	79.6	20.4	0
1801169	Slight	Slight	95.2	4.8	0	1802367	Slight	Slight	85.9	14.1	0
1801184	Slight	Slight	79	21	0	1802371	Slight	Slight	70.5	29.5	0
1801199	Slight	Slight	81.2	18.8	0						