Hindawi
Journal of Advanced Transportation
Volume 2017, Article ID 2525481, 18 pages
https://doi.org/10.1155/2017/2525481

Research Article

A Bayesian Network Approach to Causation Analysis of Road Accidents Using Netica

Xin Zou and Wen Long Yue

School of Natural and Built Environments, University of South Australia, Adelaide, SA 5095, Australia

Correspondence should be addressed to Xin Zou; zouxy011@mymail.unisa.edu.au

Received 4 June 2017; Revised 7 October 2017; Accepted 31 October 2017; Published 12 December 2017

Academic Editor: Alexandre De Barros

Copyright © 2017 Xin Zou and Wen Long Yue. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on an overall consideration of factors affecting road safety evaluations, the Bayesian network theory based on probability risk analysis was applied to the causation analysis of road accidents. By taking Adelaide Central Business District (CBD) in South Australia as a case, the Bayesian network structure was established by integrating K2 algorithm with experts' knowledge, and Expectation-Maximization algorithm that could process missing data was adopted to conduct the parameter learning in Netica, thereby establishing the Bayesian network model for the causation analysis of road accidents. Then Netica was used to carry out posterior probability reasoning, the most probable explanation, and inferential analysis. The results showed that the Bayesian network model could effectively explore the complex logical relation in road accidents and express the uncertain relation among related variables. The model not only can quantitatively predict the probability of an accident in certain road traffic condition but also can find the key reasons and the most unfavorable state combination which leads to the occurrence of an accident. The results of the study can provide theoretical support for urban road management authorities to thoroughly analyse the induction factors of road accidents and then establish basis in improving the safety performance of the urban road traffic system.

1. Introduction

With the expansion of urban development and the surging of vehicle ownership, urban travel becomes vulnerable to three "chronic diseases," which are congestion, accident, and pollution. Among the above three, accident has been recognised as the most negative aspect, in particular in and around Central Business District (CBD). According to *Global Plan for the Decade of Action for Road Safety 2011–2020* developed by the UN Road Safety Collaboration in 2011, nearly 1.3 million people die as a result of road traffic collisions per annum, which means more than 3,000 fatalities per day. And 20 to 50 million more people sustained nonfatal injuries from collisions, and these injuries were an important cause of disability worldwide. The case in Australia is also at an alarming level; there were around 25 deaths and 700 serious injuries per week, and cost to tax payers was more than 32 billion dollars a year [1]. Unless immediate and effective action is taken, road traffic injuries are predicted to become the fifth leading cause of death in the world. Therefore, the analysis and evaluation on the influencing factors on traffic accident, estimation of the potential safety hazards, and selection of appropriate measures in advance, so as to reduce the frequency and severity of traffic accidents, are important research topics in road safety engineering.

Previous studies showed that there are many reasons behind road accidents; these causes may be coherent to each other, and, for instance, poor road alignments and unexpected vehicle compositions or behaviours may result in the confusion of road users, which may lead to traffic accidents. However, many official records of road accidents indicated that most of the crashes are only pointed to single causes, especially human errors. For example, according to a crash causation survey released by the US National Highway Traffic Safety Administration (NHTSA) in 2015 [2, 3], drivers are to be criticised for 94% of crash cases. Apparently, the 94% of such accidents are also related to other causes from common experience, such as road alignment [4–6], traffic sign [7–9], and weather condition [10–12]. Therefore, the existing road accident statistics cannot fully reveal the causes, and traffic engineers and road infrastructure designers are provided with limited information for the accident mechanism and the formulation of improvement plans. It is of great importance to take full advantage of the traffic accident statistics and mine potential information so as to provide a basis for the analysis of accident mechanism and the improvement of road safety.

Bayesian network is one of the effective methods in the field of artificial intelligence to express uncertainty analysis and probability reasoning of a system. It can exploit the dependence relationships based on local conditions in a model to conduct bidirectional uncertainty investigation for prediction, classification, and diagnostic analyses. At present, there are some software platforms available for the construction of a Bayesian network, such as Bayes Net Toolbox (BNT), BayesBuilder, and JavaBayes, of which the MATLAB-based BNT developed by Murphy [13] is extensively used. This toolbox provides a lot of underlying basic function libraries for Bayesian network learning, but it does not integrate the basic functions for Bayesian network learning into a system. Moreover, BNT does not have Graphical User Interface (GUI), which is not user-friendly, nor can it be well generalized. Netica is a Bayesian network learning software developed by Norsys Software Corporation in Canada, which has been extensively applied in uncertainty management such as business, engineering, medicine, and ecology [14–16] due to its powerful functions, friendly GUI, reliable computation, and good performance. In this paper, a model is formulated using Bayesian network for road accident studies, and then a Bayesian network learning process, posterior probability reasoning, most probable explanation, and inferential analysis are conducted by using Netica.

This paper is organized as follows: Section 2 reviews the related literature on causation analysis of road accidents; Section 3 describes the construction of a Bayesian network model; Section 4 presents a case study on Bayesian network model application for Adelaide Central Business District (CBD) in South Australia; the findings of this study are summarized in Section 5.

2. Literature Review

The use of causation theory for road accident analysis aims to extract the accident mechanisms and accident models from a large number of typical accidents so as to provide theoretical basis for the qualitative and quantitative analyses, the predication and prevention of accidents, and the improvement of safety management. Scholars across the world have done some researches in road accident causation analysis and various data sources, variables, sample sizes, and analytical models, such as aggregated models which include Frequency Analysis [17–19] and χ^2 Test [20, 21].

In terms of disaggregated models, as the frequency of road accidents is in a form of nonnegative, discrete, and abnormal distribution and based on experience the frequency of accidents follows Poisson distribution, the Poisson regression model can be applied to analyse the influence of each risk factor on the frequency of accidents [22]. The negative binomial distribution regression is based on Poisson distribution, but its specification error follows Gamma distribution. The negative binomial regression model has been extensively applied in traffic safety analysis model [23–27]. However, the assumption that the mean value of Poisson distribution is equal to the variance is often inconsistent with realities. And in the analysis of longitudinal data samples, the adoptions of Poisson regression model and negative binomial regression model are likely to generate biased estimate and even incorrect results. When the explained variables only take a limited number of multiple discrete values, the established regression model is a discrete choice model, in which Logit model is the earliest discrete choice model and is one of the widely used models [28–31]. For an applicable statistical model, research object is required to be in independent distribution; while the safety data has a complex spatial distribution, the accuracy and robustness of safety level estimation will be greatly affected if the spatial feature is neglected.

Through the review of the existing literature, it has been discovered that past researches on the causation analysis of traffic accidents are gradually evolving from the descriptive simple analysis based on aggregated models to the multi-variable complex modeling analysis based on disaggregated models. However, the deficiencies of the existing studies are the following: the influencing factors on accidents are not fully considered; most are based on specific, isolated, superficial single-factor analysis, considering only the main influence factors. These studies revealed the inherent rules of the occurrence of accidents in one aspect or case but ignored the multidimensionality of accident relationships and their correlations, so that the complex logical relationship between causes, accident occurrence, and accident consequence was not reflected. Therefore, research methods and analysis technologies are not generally applicable. Although some scholars used Decision Tree [26, 32, 33], Bayesian network [34–36], and other complex systems to research the correlation between accident causes, the theoretical systems and related supporting technologies have not been systematically established.

3. Construction of Bayesian Network Model

3.1. Basic Principles of Bayesian Network. Bayesian network, also referred to as belief network, is considered as one of the most effective theoretical models in the fields of uncertainty knowledge representation and reasoning. It is a directed acyclic network topology consisting of node set and directed edge, and each node denotes one variable state, while directed edge denotes the dependence between variables. The correlation intension or confidence coefficient between variables is described by using Conditional Probability Table (CPT). Prediction, diagnosis, classification, and other tasks can be achieved by using learning and statistical inference functions of Bayes theorem. Bayesian network uses probability to denote the uncertainty of all forms and uses the probabilistic rules to achieve learning and reasoning process. It has the following relationship:

$$p(X) = \prod_{i=1}^{n} p(X_i \mid pa_i). \tag{1}$$

A set of variables $X = \{X_1, X_2, \ldots, X_n\}$ of Bayesian network consists of the following components [37] S is a network

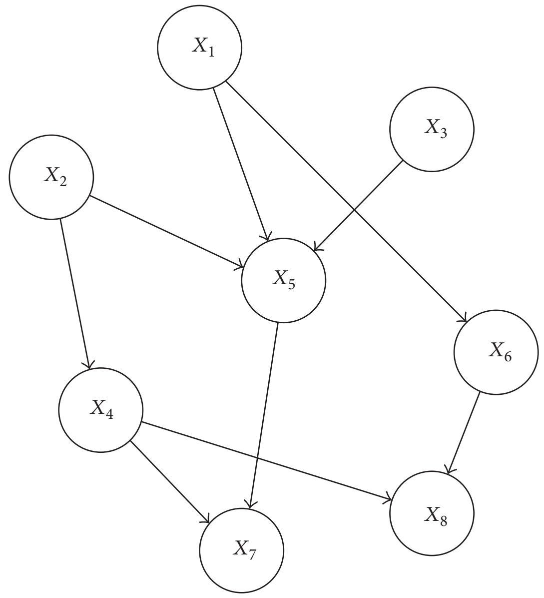

Figure 1: Graph of a valid Bayesian network (no cycle exists).

structure which denotes the conditional independent assertion in variable set X, P is a set of local probability distributions associated with each variable, X_i denotes the variable node, and pa_i denotes the father node of X_i in S.

S and P define the joint probability distribution of X. S is a directed acyclic graph (DAG), and each node in S corresponds to a variable in X (Figure 1). The default arc between nodes of S denotes conditional independence.

Use P to denote the local probability distribution in (1), namely, the product term $p(X_i \mid pa_i)$ $(i = 1, 2, \ldots, n)$; then the binary group (S, P) denotes the joint probability distribution $p(X)$.

The construction of a Bayesian network mainly involves the following steps:

(1) Structure learning: determine the factor variables (nodes) related to the study object, and then determine the dependent or independent relationship between the nodes so as to construct a directed acyclic network structure

(2) Parameter learning: based on the given Bayesian network structure, learn the Conditional Probability Table (CPT) at each node of the Bayesian network model

3.2. Structure Learning. As the network structure and data set can be used to determine the parameters, structure learning is the basis of Bayesian network learning, and the effective structure learning is the key to constructing the optimal network structure.

The construction of Bayesian network structure includes the following three points:

(1) Based on expert experience and prior knowledge, determine the variable nodes of Bayesian network so as to determine the structure of Bayesian network.

(2) Through the learning of sample data, automatically acquire the Bayesian network structure by using machine learning algorithm.

(3) Based on expert experience and machine learning of data, acquire the Bayesian network structure by using data fusion method.

As the third point combines the advantages of expert experience and machine learning and avoids the disadvantage of using one method to determine the Bayesian network structure only, in this paper, the third method to determine the Bayesian network structure for the causation analysis of road accidents will be used. Common machine learning methods include K2 algorithm, MCMC algorithm, and hill-climbing algorithm. K2 algorithm is based on the scoring function and hill-climbing algorithm, which lies in the basic principle: from an empty network, according to the predefined order of nodes, select the node with the most posterior probability as the father node of this node, sequentially traverse all nodes, and gradually add the optimal father node to each variable. K2 algorithm uses posterior probabilities as the scoring function, which is described as follows:

$$P(D \mid B_S) = \prod_{i=1}^{n} \operatorname{score}(i, pa_i), \tag{2}$$

where

$$\operatorname{score}(i, pa_i) = \prod_{j=1}^{q_i} \left[\frac{\Gamma(\partial_{ij})}{\Gamma(\partial_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\partial_{ijk} + N_{ijk})}{\Gamma(\partial_{ijk})} \right]. \tag{3}$$

D is a set of variables.

B_S is the network structure.

n are the numbers of nodes in the graph.

q_i are configurations (states) of the parents of the ith node.

r_i are mutual exclusive states of the ith node.

N_{ijk} are instances of the ith node being in the kth state when its parents are in their jth configuration, and $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

∂_{ijk} are the hyperparameters of the Dirichlet distribution and correspond to the a priori probability distribution of X_i taking on its kth state while its parents are in their jth configuration. $\partial_{ij} = \sum_{k=1}^{r_i} \partial_{ijk}$.

The gamma function $\Gamma(X) = \int_0^{+\infty} t^{X-1} e^{-t} dt$ satisfies $\Gamma(X+1) = X\Gamma(X)$ and $\Gamma(1) = 1$.

K2 algorithm uses a variable order ρ and a positive integer u to limit the search space, which seeks the optimal model $\wp$ that meets the following two conditions: (1) the number of father nodes of any variable in $\wp$ should not be greater than u and (2) ρ is a topological order of $\wp$. However, as K2 algorithm adopts greedy search strategy, which may easily fall into the local optimal solution and cannot guarantee that the network acquired is the optimal network, the knowledge and experience of experts need to be integrated so as to acquire the optimal network structure. In this paper, the combination of expert experience and K2 algorithm will perform the Bayesian network structure learning for the causation analysis of road accidents.

3.3. Parameter Learning. After determining the topological structure of Bayesian network, the parameter learning of Bayesian network can be performed. In the process of collecting road accidents information, missing data often occurs due to various reasons, for instance, recording instruments malfunction and confusion of respondents in answering questions. Most of statistical models cannot directly analyse the data with missing values, and in the case of any missing values, the record with missing values is generally eliminated directly to ensure that the statistical model can be properly fitted. If the missing values are less, this will not greatly affect the results if the record with missing values is directly eliminated. However, if the multivariate analysis is performed, more variables will be studied, which means that more records will be eliminated; it may cause a loss of information, reduce the power of test, and cause some bias to research results [38].

The Expectation-Maximization (EM) algorithm is an asymptotic deterministic estimation method for the unknown parameter θ with missing data. It can be used to perform maximum likelihood estimation on the parameters from incomplete data set, which is a practical learning algorithm [39]. EM algorithm can be widely used to deal with incomplete data, such as missing data and censored data. EM algorithm mainly involves two steps: Expectation Step (E-Step) and Maximization Step (M-Step). The algorithm is described as follows.

(1) Initialize $\theta^{(0)}$. Set accuracy ε and correction value $\widehat{\theta}'$ of estimated value $\widehat{\theta}$.

$$\begin{gathered}\text{While } \left|\widehat{\theta}-\widehat{\theta}'\right| > \varepsilon, \\ \text{do } \widehat{\theta} \longleftarrow \widehat{\theta}'.\end{gathered} \tag{4}$$

(2) E-Step. Calculate the expected sufficient statistic of missing value e^*.

The probability distribution of e^* is

$$P\left(e^* \mid e,\widehat{\theta}\right) = \frac{P\left(e \mid e^*,\widehat{\theta}\right) P\left(e^* \mid \widehat{\theta}\right)}{\sum_{e^*} P\left(e \mid e^*,\widehat{\theta}\right) P\left(e^* \mid \widehat{\theta}\right)}, \tag{5}$$

where

$$\begin{gathered}P\left(e \mid e^*,\widehat{\theta}\right) = \frac{P\left(e,e^*,\widehat{\theta}\right)}{P\left(e^*,\widehat{\theta}\right)}, \\ P\left(e^* \mid \widehat{\theta}\right) = \frac{P\left(e^*,\widehat{\theta}\right)}{P\left(\widehat{\theta}\right)}.\end{gathered} \tag{6}$$

The sufficient statistic is

$$E_{P(X|e,\widehat{\theta})} N_{ijk} = \sum_{j,k} P\left(X_{ij}, \pi\left(X_{ik}\right) \mid \widehat{\theta}\right), \tag{7}$$

where $P(e \mid e^*, \widehat{\theta})$ is the probability distribution of e under the condition that e^* and $\widehat{\theta}$ are known, $P(e^*, \widehat{\theta})$ is the joint distribution of e^* and $\widehat{\theta}$, X_i is the ith variable, N_{ijk} is the count of all possible joint instantiations between X_i and $\pi(X_i)$ denoted by j and k, respectively.

(3) M-Step. Calculate the new maximum likelihood (ML) or maximum a posteriori (MAP) values of $\widehat{\theta}'$ in the given condition $P(e^* | e, \widehat{\theta})$.

In Expectation-Maximization, we have the following:

ML:

$$\widehat{\theta}'_{ijk} = \frac{E_{P(X|e,\widehat{\theta})} N_{ijk}}{\sum_{k'} E_{P(X|e,\widehat{\theta})} N_{ijk'}}. \tag{8}$$

MAP:

$$\widehat{\theta}'_{ijk} = \frac{\alpha_{ijk} + E_{P(X|e,\widehat{\theta})} N_{ijk}}{\sum_{k'} \left(\alpha_{ijk'} + E_{P(X|e,\widehat{\theta})} N_{ijk'}\right)}, \tag{9}$$

where α_{ijk} is the Dirichlet parameter that can be obtained through the iteration process of E-Step and M-Step.

E-Step is used to calculate the expected sufficient statistic of e^*, and M-Step is used to conduct new estimation of learning parameter by using the statistic obtained in E-Step. In this paper, the Bayesian network parameter learning of road accidents is performed by using EM algorithm in Netica.

4. Case Studies

4.1. Study Area and Data Source. Adelaide Central Business District (CBD) in South Australia is selected as the study case, as it attracts 22% of metropolitan Adelaide's work trips [40] and has the first and the second most dangerous accident concentration areas which are North Terrace and West Terrace in the CBD [41].

The crash data of South Australia from 2006 to 2008 were provided by the Department of Planning, Transport and Infrastructure (DPTI), and ArcGIS 10.5 software was used to locate the precise crash sites, as shown in Figure 2.

4.2. Variable Selection and Data Preprocessing. By using ArcGIS 10.5, 1558 and 756 data sets of road accidents in Adelaide CBD from 2006-2007 and 2008 are obtained, respectively. The statistical data from 2006 to 2007 will be used for the construction of Bayesian network model and calibration, and the statistical data in 2008 will be used for the model validation process.

Previous studies [33, 42–45] provided some in-depth insights to guide the variable selection, discretization, and

FIGURE 2: The location and region of Adelaide CBD.

classification in this research. As a result, fourteen variables are selected from the data sets as having "significant influence," that is, "crash type," "driver's apparent error," "road geometry," "vehicle type," and others, as shown in Table 1. However, according to the statistical result in Table 1, it can be seen that the percentage of "inattention" reaches up to 39.84%, which is the biggest contributing factor in "driver's apparent error" category. In our daily routine, "inattention" is explained as "failure to give attention, or negligence." Generally, such usage is quite convenient for record purposes; however, from the perspective of psychology and physiology, the usage is not clear and definite. There are lots of other reasons that may be behind traffic accidents, such as human factors (driver's physical and mental state, knowledge and skill, and the operational approach), objective factors (vehicles, roads, and road facilities), and safety management. If all above factors are simply summarized as "inattention," then the causes of traffic accidents are to be extremely simplified. And the prevention measures will be hardly developed. Therefore, in this research, the factor of "inattention" will be excluded, and all variables used for modeling are shown in Table 2.

Bayesian network can be used to process continuous variables and discrete variables. As the classification result of traffic accident variables obviously has the discrete characteristic, discrete variables are adopted for Bayesian network analysis of road accidents. Before structure learning, discretization processing has to be conducted for road accident variable. The discretization values and value descriptions of processed variables are shown in Table 2.

4.3. Structure Learning. In this paper, the method combining K2 algorithm and experts' knowledge is used to formulate the Bayesian network structure. Based on K2 algorithm, FullBNT-1.0.7 is utilized to conduct structure learning via MATLAB. Through repeated selection and sequencing of variables by experts, the Bayesian network structure is finally developed, as shown in Figure 3. The network is composed by 14 nodes and several lines. The 14 nodes refer to 14 variables, and lines between these nodes indicate the relationships among the variables.

It can be seen from Figure 3 that some road accident variables have demonstrated clear hierarchical relations of affecting others and being affected by others. Road accidents result from the interaction of variables from "traffic participant, vehicle, road, and environment," which is fully reflected by the Bayesian network structure as well. For instance, "vehicle movement" is affected by "road geometry" and "driver's apparent error," but it can also affect "total units involved" at the same time. The actual situation of road accidents can be fully embodied by the interactional hierarchical relationship of variables in Bayesian network.

4.4. Parameter Learning. Once a Bayesian network structure is formed, parameter learning can be carried out. The Bayesian network structure can be created in Netica, and then parameter learning can be conducted, thus obtaining the conditional probability distribution of nodes. Finally, the Bayesian network model for the road accident causation analysis can be determined, as shown in Figure 4.

4.5. Model Calibration and Validation. To validate the parameter learning accuracy and prediction accuracy of the Bayesian network model, sensitivity analysis is used to identify the sensitive factors with a significant impact on the target node from a number of uncertain factors, and then the target node is set as the evidence variable to conduct model fitting and prediction with these sensitive factors.

TABLE 1: Variables of road accidents, Adelaide CBD, 2006–2008.

Variable class	Variable name	Discretization value	Value description	Frequency	Percentage
Driver	Apparent error (X_1)	1	Fail to stand	307	13.25%
		2	Change lanes to endanger	221	9.54%
		3	Incorrect turn	31	1.34%
		4	Reverse without due care	92	3.97%
		5	Follow too closely	173	7.47%
		6	Overtake without due care	52	2.24%
		7	Disobey traffic lights	171	7.38%
		8	Disobey stop sign	23	0.99%
		9	Disobey give way sign	47	2.03%
		10	Inattention	923	39.84%
		11	DUI	23	0.99%
		12	Fail to give way	254	10.96%
Road	Road geometry (X_2)	1	Cross road	1116	48.17%
		2	Y junction	57	2.46%
		3	T junction	450	19.42%
		4	Multiple	33	1.42%
		5	Divided road	349	15.06%
		6	Not divided	294	12.69%
		7	Pedestrian crossing	18	0.78%
	Road moisture condition (X_3)	1	Wet	237	10.23%
		2	Dry	2080	89.77%
	Traffic control (X_4)	1	Traffic signals	1282	55.33%
		2	Stop sign	42	1.81%
		3	Give way sign	118	5.09%
		4	No control	875	37.76%
Environment	Weather condition (X_5)	1	Raining	151	6.52%
		2	Not raining	2166	93.48%
	Light condition (X_6)	1	Daylight	1716	74.06%
		2	Night	601	25.94%
Vehicle	Vehicle type (X_7)	1	Heavy	172	7.42%
		2	Medium	901	38.89%
		3	Light	1244	53.69%
	Vehicle movement (X_8)	1	Right turn	426	18.39%
		2	Left turn	95	4.10%
		3	U turn	122	5.27%
		4	Swerving	242	10.44%
		5	Reversing	77	3.32%
		6	Straight ahead	1238	53.43%
		7	Entering private driveway	15	0.65%
		8	Leaving private driveway	50	2.16%
		9	Overtaking on right	39	1.68%
		10	Overtaking on left	13	0.56%

Table 1: Continued.

Variable class	Variable name	Discretization value	Value description	Frequency	Percentage
Road crash	Crash type (Y_1)	1	Rear end	1010	43.59%
		2	Hit fixed object	59	2.55%
		3	Side swipe	392	16.92%
		4	Right angle	377	16.27%
		5	Head on	5	0.22%
		6	Hit pedestrian	50	2.16%
		7	Right turn	333	14.37%
		8	Hit parked vehicle	91	3.93%
	Crash severity (Y_2)	1	PDO (property damage only)	1748	75.44%
		2	Injury	569	24.56%
	Total units (involved in a road crash) (Y_3)	1	Two units	2012	86.84%
		2	Three units	258	11.14%
		3	Four units	40	1.73%
		4	Five units	7	0.30%
	Total casualties (fatalities and treated injuries) (Y_4)	1	None	1748	75.44%
		2	One casualty	495	21.36%
		3	Two casualties	62	2.68%
		4	Three casualties	12	0.52%
	Total serious injuries (Y_5)	1	None	2267	97.84%
		2	One serious injury	50	2.16%
	Total estimated damage (A$) ($Y_6$)	1	[0, 5000)	1139	49.16%
		2	[5000, 10000)	795	34.31%
		3	[10000, +∞)	383	16.53%

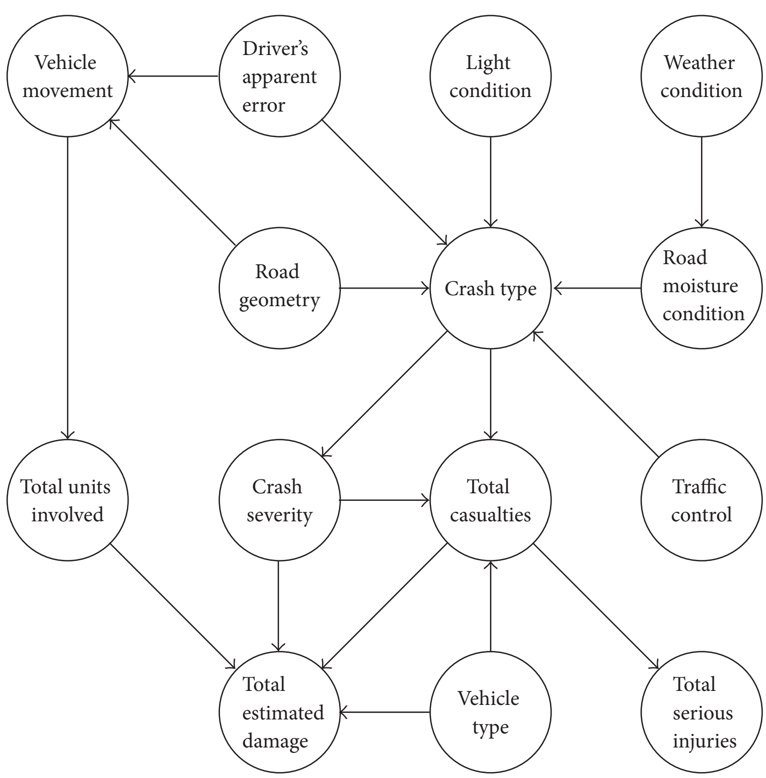

Figure 3: The Bayesian network structure for the road accident analysis.

Table 2: Variables used for the Construction of Bayesian Network.

Variable class	Variable name	Discretization value	Value description	Frequency	Percentage
Driver	Apparent error (X_1)	1	Fail to stand	199	21.47%
		2	Change lanes to endanger	141	15.21%
		3	Incorrect turn	24	2.59%
		4	Reverse without due care	61	6.58%
		5	Follow too closely	119	12.84%
		6	Overtake without due care	35	3.78%
		7	Disobey traffic lights	115	12.41%
		8	Disobey stop sign	18	1.94%
		9	Disobey give way sign	27	2.91%
		10	DUI	9	0.97%
		11	Fail to give way	179	19.31%
Road	Road geometry (X_2)	1	Cross road	454	48.98%
		2	Y junction	11	1.19%
		3	T junction	196	21.14%
		4	Multiple	13	1.40%
		5	Divided road	132	14.24%
		6	Not divided	121	13.05%
	Road moisture condition (X_3)	1	Wet	100	10.79%
		2	Dry	827	89.21%
	Traffic control (X_4)	1	Traffic signals	477	51.46%
		2	Stop sign	27	2.91%
		3	Give way sign	56	6.04%
		4	No control	367	39.59%
Environment	Weather condition (X_5)	1	Raining	65	7.01%
		2	Not raining	862	92.99%
	Light condition (X_6)	1	Daylight	677	73.03%
		2	Night	250	26.97%
Vehicle	Vehicle type (X_7)	1	Heavy	63	6.80%
		2	Medium	379	40.88%
		3	Light	485	52.32%
	Vehicle movement (X_8)	1	Right turn	267	28.80%
		2	Left turn	41	4.42%
		3	U turn	79	8.52%
		4	Swerving	137	14.78%
		5	Reversing	52	5.61%
		6	Straight ahead	275	29.67%
		7	Entering private driveway	10	1.08%
		8	Leaving private driveway	31	3.34%
		9	Overtaking on right	25	2.70%
		10	Overtaking on left	10	1.08%

Table 2: Continued.

Variable class	Variable name	Discretization value	Value description	Frequency	Percentage
Road crash	Crash type (Y_1)	1	Rear end	155	16.72%
		2	Hit fixed object	4	0.43%
		3	Side swipe	255	27.51%
		4	Right angle	258	27.83%
		5	Head on	3	0.32%
		6	Hit pedestrian	22	2.37%
		7	Right turn	216	23.30%
		8	Hit parked vehicle	14	1.51%
	Crash severity (Y_2)	1	PDO (property damage only)	689	74.33%
		2	Injury	238	25.67%
	Total units (involved in a road crash) (Y_3)	1	Two units	860	92.77%
		2	Three units	54	5.83%
		3	Four units	9	0.97%
		4	Five units	4	0.43%
	Total casualties (fatalities and treated injuries) (Y_4)	1	None	689	74.33%
		2	One casualty	213	22.98%
		3	Two casualties	21	2.27%
		4	Three casualties	4	0.43%
	Total serious injuries (Y_5)	1	None	904	97.52%
		2	One serious injury	23	2.48%
	Total estimated damage (A\$) (Y_6)	1	[0, 5000)	423	45.63%
		2	[5000, 10000)	342	36.89%
		3	[10000, +∞)	162	17.48%

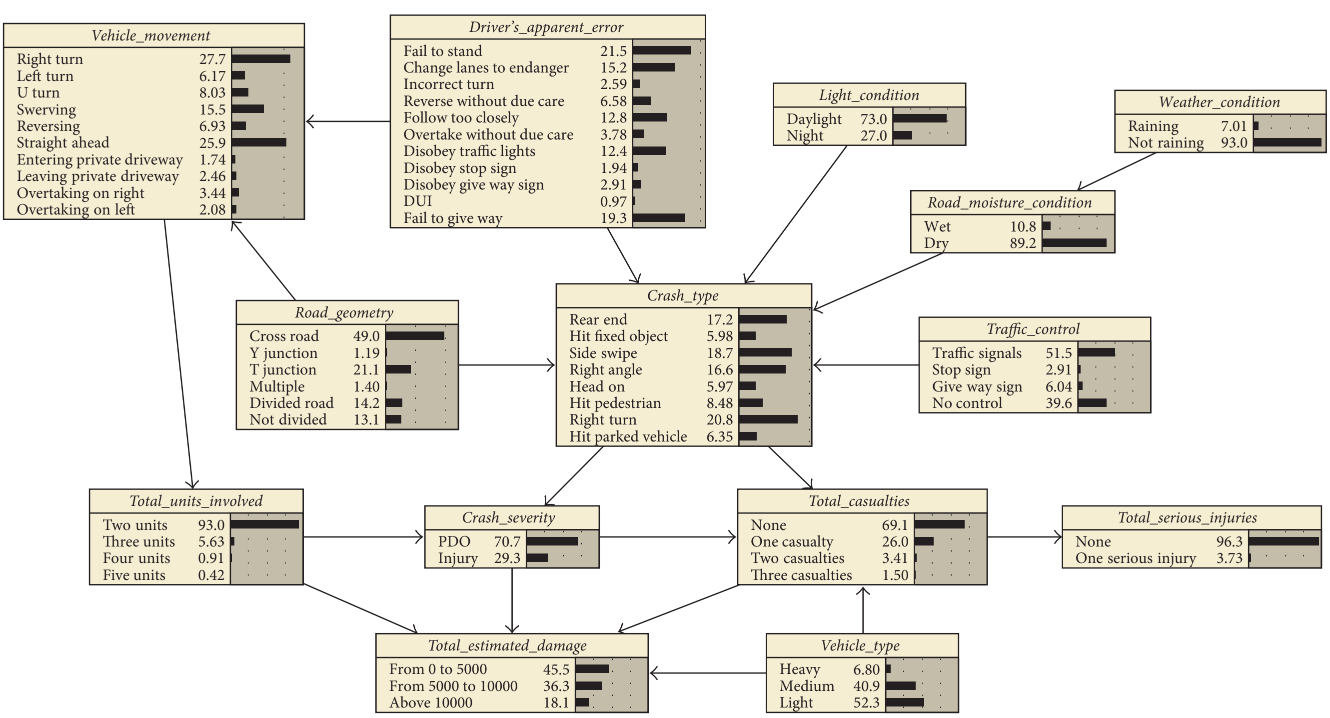

Figure 4: The Bayesian network model after parameter learning in Netica 6.02.

TABLE 3: Sensitivity analysis result of the node "crash type."

Node	Mutual info	Percent	Variance of beliefs
Crash type	2.82978	100	0.7195390
Driver's apparent error	0.57217	20.2	0.0869963
Vehicle movement	0.33948	12	0.0351281
Total casualties	0.21563	7.62	0.0057073
Crash severity	0.17897	6.32	0.0037534
Road geometry	0.04044	1.43	0.0008944
Total serious injuries	0.01888	0.667	0.0004707
Traffic control	0.01125	0.398	0.0001573
Road moisture condition	0.00503	0.178	0.0001027
Light condition	0.00417	0.147	0.0000915
Total estimated damage	0.00407	0.144	0.0000533
Total units involved	0.00394	0.139	0.0002397
Weather condition	0.00313	0.111	0.0000638
Vehicle type	0.00000	0.000	0.0000000

4.5.1. Sensitivity Analysis. In Bayesian network, the sensitivity analysis refers to the analysis of the influence and influence degrees of multiple causes (node states) on result (target node). Based on sensitivity analysis, the elementary events with relatively greater contribution to the probabilities of the consequential events can be determined to facilitate the reduction of probabilities of these elementary events by taking effective measures, so that the probabilities of the consequential events will be reduced.

The sensitivity analysis function of Netica can be used to identify which factors have more important safety management values in analyzing traffic accidents. In Netica, select the target node and then analyse the impact degrees of other nodes on the target node in a descending order. Taking the node "crash type" as an example, make sensitivity analysis of it and the result is as shown in Table 3.

The mutual information refers to the direct or indirect information flow rate and measures the degree of dependence between nodes. In other words, the mutual information between two nodes can indicate if the two nodes are dependent on each other and if so how close their relationship is [46]. As shown in Table 3, it can be seen that the mutual info (=0.57217) of node "driver's apparent error" is the largest, which means that it has the strongest impact on "crash type," followed by "vehicle movement" and "road geometry" which have mutual info = 0.33948 and 0.04044, respectively.

4.5.2. Model Fitting and Prediction. Based on the sensitivity analysis result of "crash type," the posterior probabilities of "driver's apparent error," "vehicle movement," and "road geometry" obtained from the Bayesian network are compared with the actual calculations from 2006 to 2007 and from 2008, respectively. Due to the large amount of data, "rear end" from the parameter learning results of "crash type" is used for exemplificative explanation.

Figures 5, 6, and 7 show the posterior and actual probability distributions of "driver's apparent error," "vehicle

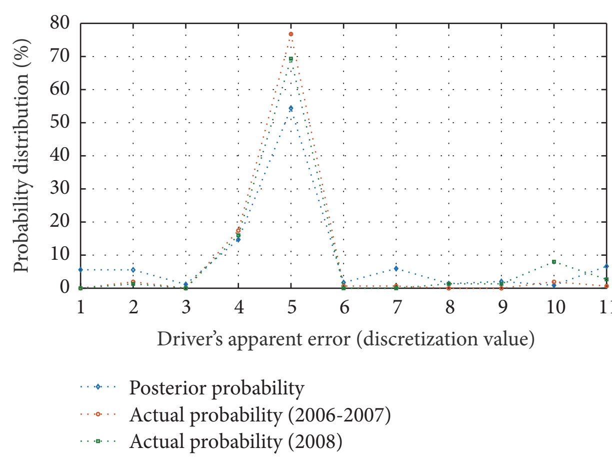

FIGURE 5: The comparison of posterior and actual probability curves of driver's apparent error when the evidence variable is "rear end."

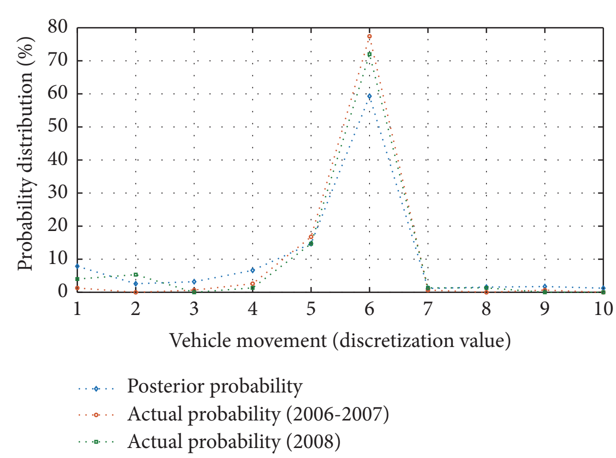

FIGURE 6: The comparison of posterior and actual probability curves of vehicle movement when the evidence variable is "rear end."

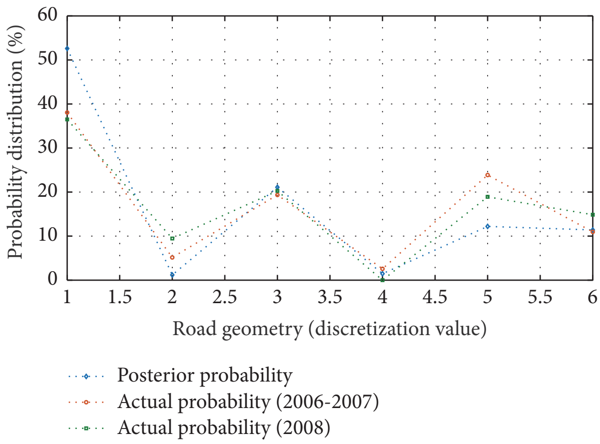

Figure 7: The comparison of posterior and actual probability curves of road geometry when the evidence variable is "rear end."

movement," and "road geometry," respectively, when the evidence variable is "rear end." Compared with the actual computational results from 2006 to 2007, the maximum mean absolute error (MAE) of Bayesian network model is 5.58%. Similarly, compared with the actual computational results of 2008, the maximum MAE is 6.14%, which suggests that the Bayesian network model has both high fitting accuracy and high prediction accuracy. Therefore, it is feasible to use the Bayesian network model to conduct result prediction and inferential analysis of each variable of road accidents accordingly.

4.6. Bayesian Network Model Application

4.6.1. Posterior Probability Reasoning. Bayesian network model can be used to perform probability reasoning, including posterior probability calculations. Precisely, it aims to calculate the posterior probabilities of some targeted nodes, control the influence degrees of determined specific nodes on the nodes of interest, predict the possibility of accident occurrence, and analyse the major accident sources under the condition that states of specific nodes are determined. In brief, the posterior probabilities in inferring the result from cause and inferring the cause from result are referred to as accident prediction and causal inference, respectively.

(1) Accident Prediction. Figure 8 is the accident prediction on the assumption that a driver "disobeys traffic lights" when driving in Adelaide CBD. Input the evidence variables ("traffic signals" and "disobey traffic lights") emerging from this circumstance into the Bayesian network, so it becomes a problem to solve the posterior probabilities of other nodes, with the known status of some evidence variables.

In Netica, set both the statuses of "traffic signals" and "disobey traffic lights" as 100%; that is, the statuses of the evidence variables are determined. Then update the probabilities of the whole network; the probability change of relevant nodes, namely, the probability change of "crash type" and other nodes, can be observed. In this case, the probability of "right angle" in "crash type" is found to increase from the initial 16.6% to 58.0%. This suggests that if the driver "disobeys traffic lights," the probability of "right angle" will significantly increase.

As shown in Figure 9, in addition to "disobey traffic lights," assume that the driving time is at night; namely, the status of "night" in "light condition" is set as 100%. After automatically updating the probabilities of the whole network, the probability of "right angle" is found to further increase from 58.0% to 59.4%, which means that the probability of "right-angle" traffic accident is higher. Go one step further and assume that it is also a rainy night (namely, set the status of "raining" in "weather condition" as 100%). According to Figure 10, once again, it can be found that the probability of "right angle" further increases from 59.4% to 63.3%. This suggests that "driver's apparent error," "light condition," and "weather condition" will all affect the probability of "right angle" to various degrees. Therefore, it can be found that the status change of evidence node variables will affect the probabilities of query nodes, which is consistent with the engineering practice.

(2) Causal Inference. Another important application of the Bayesian network is fault diagnosis of the system. The bidirectional reasoning technology of the Bayesian network can calculate not only the probability of a system failure under combined fault conditions but also the posterior probabilities of various components under the system fault condition and easily find out the most likely combination that caused system failure, thereby making the computational analysis more intuitive and flexible.

Conduct causal inference by taking the "side swipe" in "crash type" as an example. In this case, the evidence variable is "side swipe," so its status probability is 100%. As shown in Figure 11, after inputting the evidence, the probability of "change lanes to endanger" in "driver's apparent error" increases greatly from 15.2% to 46.7% through the automatic updating function of Netica. And the probability of "swerving" in "vehicle movement" also increases from 15.5% to 44.9%, which reaches the maximum probability. This suggests that, in the absence of other evidences, the most probable cause to "side swipe" is "swerving" (vehicle) caused by "change lanes to endanger" (driver).

4.6.2. Most Probable Explanation. Bayesian network model can be used to make the most probable explanations, precisely, from sets of multiple causes (node states) which are likely to lead to a conclusion; use Netica to identify the set that is most likely to lead to the result, and this set with the maximum likelihood will be the most probable explanation.

In the example of "side swipe" as illustrated in Figure 12, use "Most Probable Explanation" function in Netica to find out the most probable cause set. As shown in Figure 12, the most probable explanation cause (node state) set of "side swipe" is [cross road, change lanes to endanger, swerving, daylight, not raining, dry, traffic signals]. It explicitly shows that most probable explanation and causal inference are highly consistent when the evidence variable is "side swipe," and the set is also consistent with the engineering practice.

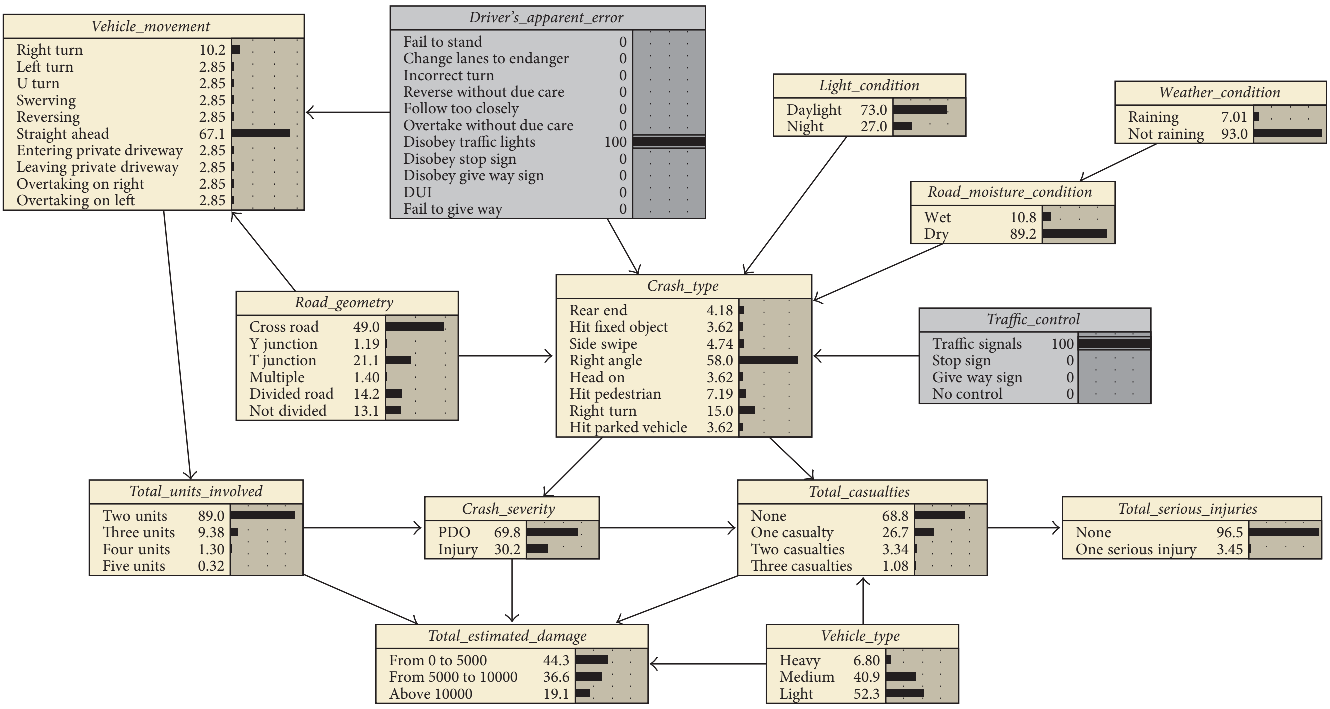

FIGURE 8: Road accident prediction when the evidence variables are "traffic signals" and "disobey traffic lights."

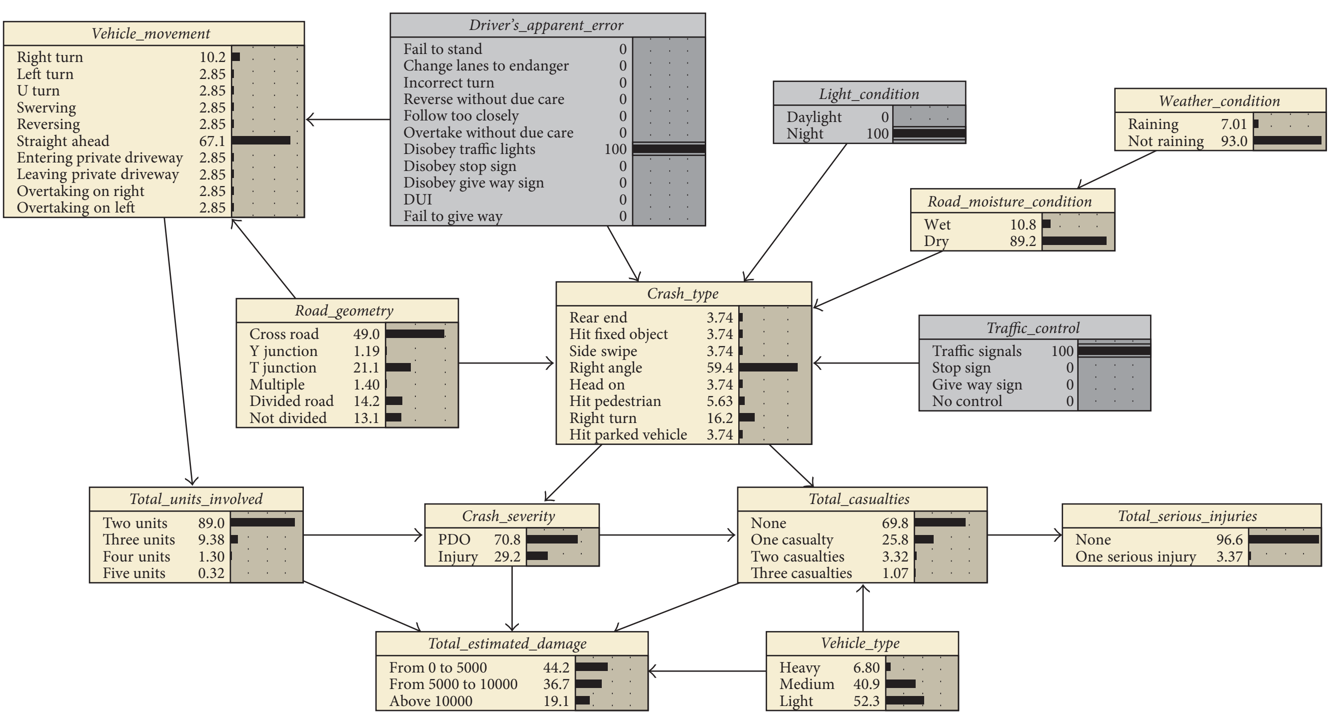

FIGURE 9: Road accident prediction when the evidence variables are "traffic signals," "disobey traffic lights," and "night."

4.7. Inferential Analysis of Accidents Based on "Serious Injuries" and "Total Estimated Damage". The application of the Bayesian network model in Netica to solve the posterior probability reasoning problem, maximum posterior hypothesis problem, and most probable explanation problem highlighted the inferential capability of the Bayesian network model. To further analyse the factors contributing to traffic accidents, especially serious traffic accidents, the Bayesian network model was used to calculate the probabilities of "serious injuries" and "total estimated damage over 10,000

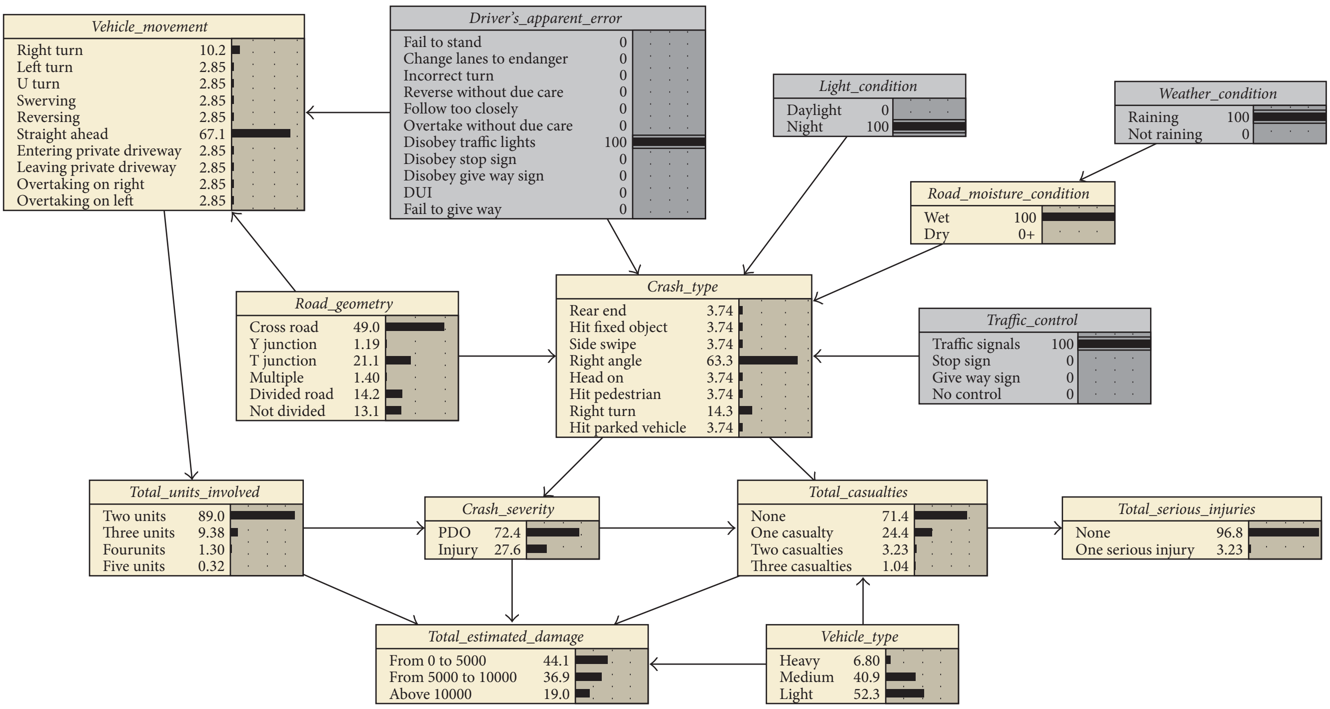

Figure 10: Road accident prediction when the evidence variables are "traffic signals," "disobey traffic lights," "night," and "raining."

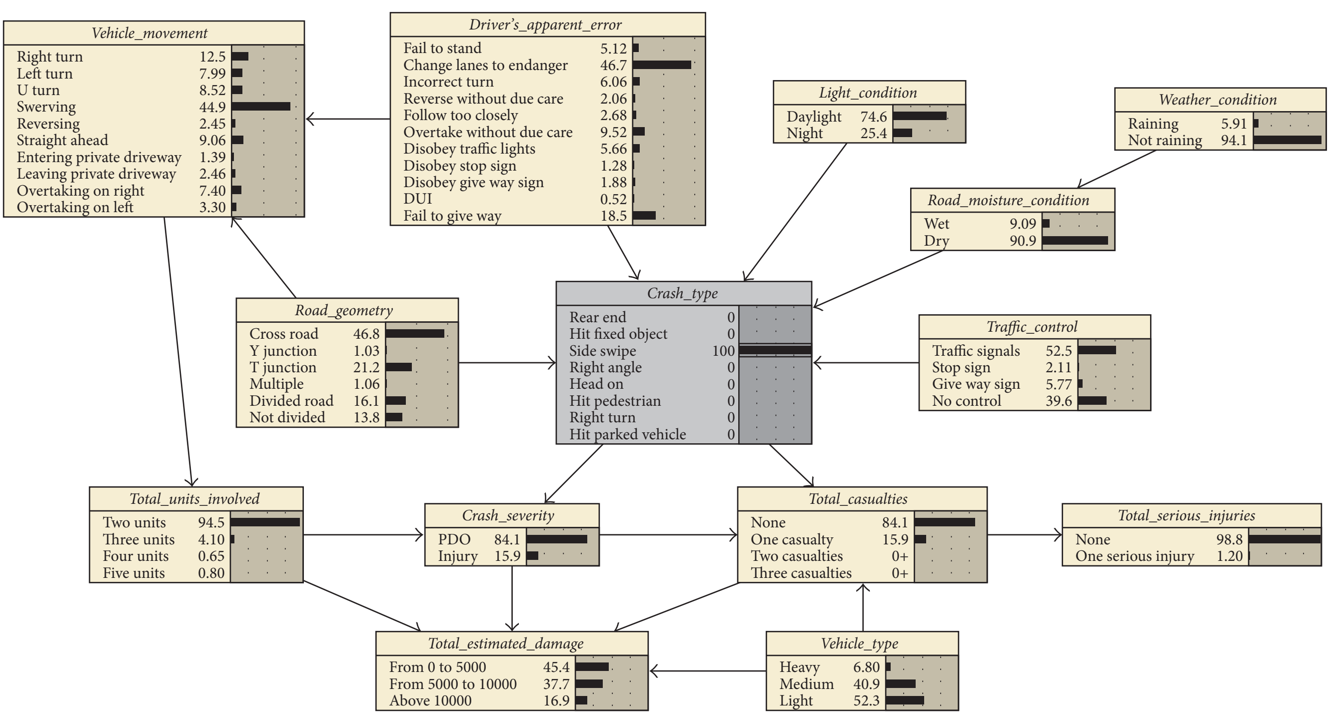

Figure 11: The posterior probability when the evidence variable is "side swipe."

AUD" under the influence of "driver's apparent error," "road geometry," "weather condition," "light condition," and "crash type," respectively. The results are shown in Table 4.

4.7.1. Driver's Apparent Error. The results presented in Table 4 indicate that "driving under the influence" (DUI) will most likely cause serious injuries and heavy property damage, as DUI is more easily to lead to dangerous behaviours including speeding, not wearing a safety belt, and reckless or erratic driving. According to the inference results, DUI is most likely to cause traffic accidents on cross roads with the inference probability of 49.0%. As for the crash types, the inference

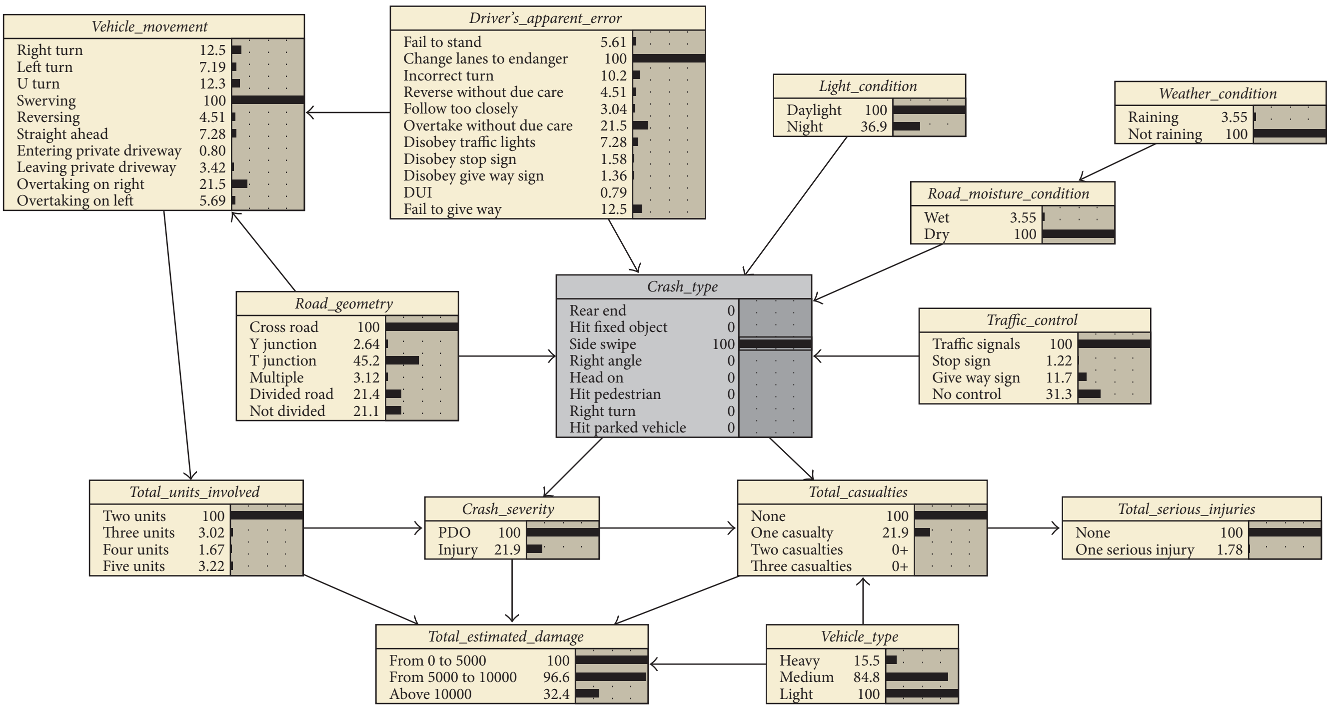

Figure 12: The most probable explanation when the evidence variable is "side swipe."

probabilities of rear-end and head-on crashes are the two highest (resp., 15.6% and 15.4%). In previous studies, among all traffic accidents caused by DUI, unrestrained occupants were 4.70 times more likely to die or 4.66 times more likely to be injured than restrained occupants [47]. Besides, drunk drivers show weak control ability of vehicles, and the higher ethanol content in their blood, the higher probability that they will have illegal actions mentioned previously [48], which are more likely to cause serious traffic accidents with heavy casualties and property damage.

4.7.2. Road Geometry. Intersections are an important part of road system and are potentially the most dangerous locations in a network as well. Previous studies have shown that intersections, especially cross roads, have higher crash rates and greater crash severity, particularly in urban areas [49–51]. As shown in Table 4, the most dangerous "road geometry" in Adelaide CBD also is "cross road," which means that the probabilities of causing serious injuries and heavy property damage at "cross roads" are both the biggest. According to the accident records in Adelaide CBD, among serious traffic accidents caused at cross roads, there are two main crash types, "right angle" and "right turn," respectively, accounting for 44.35% and 42.61%, while "fail to stand" and "disobey traffic lights" become the two main reasons of traffic accidents caused at cross roads, respectively, accounting for 38.26% and 34.78%.

4.7.3. Weather Condition. Rainfall will not only decrease the effectiveness of drivers' visual search [12] but also lower the friction coefficient of roads, which makes roads slippery, increases braking distance greatly, and thus results in the possibility of traffic accidents. A study for Melbourne, Australia, by Keay and Simmonds [10] found that rainfall was the strongest factor that correlated to weather parameter and it had the greatest impact in winter and spring. Keay and Simmonds [52] also found a contributing parameter, which is the lagged effect of rain. Symons and Perry [53] found that wet roads or raining is increasing the probability of traffic accidents which can reach up to 70 percent. Similarly, Qiu and Nixon [11] found that rain can increase the crash rate by 71% and the injury rate by 49%. This coincides with the results found in this research, which indicate that rainfall is associated with traffic accidents that had serious injuries and heavy property damage.

4.7.4. Light Condition (Urban Heat Island). Traditional bituminous pavement can absorb and store large amounts of heat during the day and continuously output the heat to the external environment at night, which will result in an increase in external environment temperature and lead to urban heat island (UHI). UHI might have some impacts on road durability and safety, such as accelerating bituminous pavement aging and exacerbating road high-temperature rutting may lead to road accidents. With the continuous expansion of city size, the comprehensive phenomenon of such microclimatic variation will become increasingly obvious [54]. As one of the major Australian cities, Adelaide is also affected by UHI [55]. And as waste heat from vehicles and temperature regulation of buildings is an important determinant of UHI magnitudes, Adelaide CBD, which has the largest density of traffic network and the largest number of buildings, has become the center of UHI. As shown in Table 4, the number of road accidents that occurred at night demonstrates a larger proportion of

Table 4: Inference results for variables that are associated with "serious injuries" and "total estimated damage" in serious traffic accidents.

Variable class	Variable name	Serious injuries/%	Total estimated damage (≥10,000 AUD)/%
Driver's apparent error	Fail to stand	3.54	17.5
	Change lanes to endanger	2.83	17.5
	Incorrect turn	3.61	17.1
	Reverse without due care	4.03	17.9
	Follow too closely	3.91	20.0
	Overtake without due care	3.36	18.7
	Disobey traffic lights	4.15	19.4
	Disobey stop sign	4.81	19.5
	Disobey give way sign	4.68	18.9
	DUI	5.17	20.4
	Fail to give way	3.91	17.2
Road geometry	Cross road	4.47	19.6
	Y junction	3.69	18.2
	T junction	3.27	18.0
	Multiple	4.17	18.4
	Divided road	4.05	18.1
	Not divided	4.12	18.3
Weather condition	Raining	4.22	18.4
	Not raining	3.69	18.1
Light condition	Daylight	3.65	18.1
	Night	3.93	18.2
Crash type	Rear end	3.40	19.0
	Hit fixed object	2.76	17.4
	Side swipe	1.20	16.9
	Right angle	2.48	17.6
	Head on	15.4	26.0
	Hit pedestrian	9.37	18.6
	Right turn	2.93	17.2
	Hit parked vehicle	0.34	16.9

55.56% than by daylight. Under the influence of UHI, Jusuf et al. [56] found that, at nighttime, commercial area has the highest ambient temperature among the four land use types (commercial, residential, industrial, and airport). Similarly, Parker [57] also found that the urban heat island is strongest at night in high-rise city centers. Therefore, the correlation analysis between road accidents and nighttime UHI is a potential research direction.

4.7.5. Crash Type. Head-on crashes are among the most severe collision types and are of great concern to road safety authorities [58]. For instance, according to an annual report presented by NHTSA in 2015, head-on crashes occupied only 2.3% of total crashes; however, they accounted for 9.6% of fatal crashes. As shown in Table 4, no matter "serious injuries" or "property damage over 10,000 AUD," the possibility of "head-on" crashes always takes the first place and is largely higher than other types of traffic accidents. As for the road geometry, the head-on crashes are most likely to happen on cross roads with the inference probability of 39.6%, while the inference probability of "disobey traffic lights" is the biggest in drivers' apparent errors which cause head-on crashes. These results agree with Bham et al. [59] who found that head-on collisions were at a higher risk for severe injuries compared with other collision types. And Rizzi et al. [60] concluded that 31% and 21% of crashes at intersections could have been avoided entirely or influenced by anti-lock braking system (ABS); however, the head-on crashes were the only crash type for which ABS seemed to be ineffective.

5. Conclusion and Future Work

As road accidents are unexpected, random, complex, and latent, it is necessary to conduct investigations on accident mechanism and accurately identify the exact causes. The Bayesian network combining the probability theory with graph theory not only has a rigorous mathematical consistency but also has the structure chart that can intuitively identify problems. Therefore, it is one of the most powerful and effective tools to deal with uncertainties.

The occurrence of road accident results from the interactions of "traffic participant, vehicle, road, and environment," and there is a potential hierarchical relation (impacting and impacted) among the variables. In Bayesian network, the directed acyclic graph is a visual expression form that is closer to the characteristics of thought and reasoning mode of human. In the study, the Bayesian network structure for the road accident causation analysis was achieved by using K2 algorithm and experts' knowledge which combines the advantages of machine learning and experts' knowledge. The structure learning result of the Bayesian network fully reflects the hierarchical relations among the accident related variables and allows for better prediction and analysis of the characteristics of road accidents.

In this study, the Bayesian network model for the road accident causation analyses was established by using Netica, Bayesian network-based software with friendly GUI. The Expectation-Maximization algorithm that can deal with missing data was adopted to process the parameter learning, and then the calibration and validation, posterior probability reasoning, most probable explanation, and inferential analysis were carried out after the construction of the Bayesian network model. The results showed that the Bayesian network model is feasible and effective for road accident causation analyses; in particular the use of posterior probability of the Bayesian network can not only more precisely and quickly find the key causes for traffic accidents but also identify the most likely cause (state) combination. The result can be used as an important theoretical basis in developing road traffic management strategies so as to improve road traffic safety.

Follow-up studies will consider the rationality of the Bayesian model and other factors that may lead to traffic accidents and establish a more accurate and comprehensive model. As the Bayesian model is a probabilistic model, more comprehensive and extensive basic data are needed to enhance its reliability, in which some data can be obtained only by carrying out experiments, despite the ability of the Bayesian mode to make up for missing data. Moreover, as the influencing road accident factors in reality are more than the factors used in this study and as Netica is also suitable for the establishment of a larger and more complex accident analysis model, the model can be expanded to a more sophisticated model that can consider more factors.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Australasian College of Road Safety, 2017 ACRS Submission to Federal Parliamentarians, Retrieved from http://acrs.org.au/wp-content/uploads/2017-ACRS-Submission-to-Federal-Parliamentarians-FINAL.pdf.

[2] National Highway Traffic Safety Administration (NHTSA)., "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," vol. 85, 101 pages, 2015.

[3] National Highway Traffic Safety Administration (NHTSA), "Traffic safety facts 2014: a compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system," 2015.

[4] M. G. Karlaftis and I. Golias, "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accident Analysis Prevention*, vol. 34, no. 3, pp. 357–365, 2002.

[5] R. Fu, Y. Guo, W. Yuan, H. Feng, and Y. Ma, "The correlation between gradients of descending roads and accident rates," *Safety Science*, vol. 49, no. 3, pp. 416–423, 2011.

[6] J. P. Gooch, V. V. Gayah, and E. T. Donnell, "Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads," *Accident Analysis Prevention*, vol. 92, pp. 71–81, 2016.

[7] H. Al-Madani and A. R. Al-Janahi, "Role of drivers personal characteristics in understanding traffic sign symbols," *Accident Analysis Prevention*, vol. 34, no. 2, pp. 185–196, 2002.

[8] D. de Waard, F. J. J. M. Steyvers, and K. A. Brookhuis, "How much visual road information is needed to drive safely and comfortably?" *Safety Science*, vol. 42, no. 7, pp. 639–655, 2004.

[9] T. Ben-Bassat and D. Shinar, *Ergonomic Guidelines for Traffic Sign Design Increase Sign Comprehension. Human Factors*, vol. 48, no. 1, pp. 182–195, 2006.

[10] K. Keay and I. Simmonds, "The association of rainfall and other weather variables with road traffic volume in melbourne, australia," *Accident Analysis Prevention*, vol. 37, no. 1, pp. 109–124, 2005.

[11] L. Qiu and W. Nixon, "Effects of adverse weather on traffic crashes: systematic review and meta-analysis," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 139–146, 2008.

[12] P. Konstantopoulos, P. Chapman, and D. Crundall, "Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 827–834, 2010.

[13] K. P. Murphy, "The bayes net toolbox for MATLAB," *Computing Science and Statistics*, vol. 33, no. 2, pp. 1024–1034, 2001, http://www.interfacesymposia.org/I01/I2001Proceedings/KMurphy/KMurphy.pdf.

[14] O. Arsene, I. Dumitrache, and I. Mihu, "Medicine expert system dynamic Bayesian Network and ontology based," *Expert Systems with Applications*, vol. 38, no. 12, 2011.

[15] G. Büyüközkan, G. Kayakutlu, and İ. S. Karakadılar, "Assessment of lean manufacturing effect on business performance using bayesian belief networks," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6539–6551, 2015.

[16] H. Mcheick, H. Nasser, M. Dbouk, and A. Nasser, "Stroke prediction context-aware health care system," in *Proceedings of IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 30–35, Washington, Wash, USA, 2016.

[17] S. Jensen, "Pedestrian Safety in Denmark," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1674, pp. 61–69, 1999.

[18] M. Stone and J. Broughton, "Getting off your bike: cycling accidents in great britain in 1990–1999," *Accident Analysis Prevention*, vol. 35, no. 4, pp. 549–556, 2003.

[19] D. E. Lefler and H. C. Gabler, "The fatality and injury risk of light truck impacts with pedestrians in the United States," *Accident Analysis Prevention*, vol. 36, no. 2, pp. 295–304, 2004.

[20] O. T. Holubowycz, "Age, sex, and blood alcohol concentration of killed and injured pedestrians," *Accident Analysis & Prevention*, vol. 27, no. 3, pp. 417–422, 1995.

[21] A. S. Al-Ghamdi, "Pedestrian–vehicle crashes and analytical techniques for stratified contingency tables," *Accident Analysis Prevention*, vol. 34, no. 2, pp. 205–214, 2002.

[22] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle

crashes: balancing statistical fit and theory," *Accident Analysis & Prevention*, vol. 37, no. 1, pp. 35–46, 2005.

[23] M. Poch and F. Mannering, "Negative binomial analysis of intersection-accident frequencies," *Journal of Transportation Engineering*, vol. 122, no. 2, pp. 105–113, 1996.

[24] J.-L. Martin, "Relationship between crash rate and hourly traffic flow on interurban motorways," *Accident Analysis & Prevention*, vol. 34, no. 5, pp. 619–629, 2002.

[25] K.-s. Ng, W.-t. Hung, and W.-g. Wong, "An algorithm for assessing the risk of traffic accident," *Journal of Safety Research*, vol. 33, no. 3, pp. 387–410, 2002.

[26] L.-Y. Chang and W.-C. Chen, "Data mining of tree-based models to analyze freeway accident frequency," *Journal of Safety Research*, vol. 36, no. 4, pp. 365–375, 2005.

[27] G. Lovegrove and T. Sayed, "Macrolevel collision prediction models to enhance traditional reactive road safety improvement programs," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 65–73, 2007.

[28] M. J. M. Sullman, M. L. Meadows, and K. B. Pajo, "Aberrant driving behaviours amongst new zealand truck drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 5, no. 3, pp. 217–232, 2002.

[29] J.-K. Kim, S. Kim, G. F. Ulfarsson, and L. A. Porrello, "Bicyclist injury severities in bicyclemotor vehicle accidents," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 238–251, 2007.

[30] P. Savolainen and F. Mannering, "Probabilistic models of motorcyclists injury severities in single- and multi-vehicle crashes," *Accident Analysis & Prevention*, vol. 39, no. 5, pp. 955–963, 2007.

[31] R. K. Young and J. Liesman, "Estimating the relationship between measured wind speed and overturning truck crashes using a binary logit model," *Accident Analysis & Prevention*, vol. 39, no. 3, pp. 574–580, 2007.

[32] P. M. Kuhnert, K.-A. Do, and R. McClure, "Combining non-parametric models with logistic regression: an application to motor vehicle injury data," *Computational Statistics & Data Analysis*, vol. 34, no. 3, pp. 371–386, 2000.

[33] J. Abellán, G. López, and J. de Oña, "Analysis of traffic accident severity using decision rules via decision trees," in *Expert Systems with Applications*, vol. 40, pp. 6047–6054, 2013.

[34] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on spanish rural highways using bayesian networks," *Accident Analysis Prevention*, vol. 43, no. 1, pp. 402–411, 2011.

[35] S. Heydari, L. F. Miranda-Moreno, D. Lord, and L. Fu, "Bayesian methodology to estimate and update safety performance functions under limited data conditions: A sensitivity analysis," *Accident Analysis Prevention*, vol. 64, pp. 41–51, 2014.

[36] A. C. Mbakwe, A. A. Saka, K. Choi, and Y.-J. Lee, "Alternative method of highway traffic safety analysis for developing countries using delphi technique and Bayesian network," *Accident Analysis & Prevention*, vol. 93, pp. 135–146, 2016.

[37] K. G. Olesen and A. L. Madsen, "Maximal prime subgraph decomposition of Bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 32, no. 1, pp. 21–31, 2002.

[38] S. Demissie, M. P. LaValley, N. J. Horton, R. J. Glynn, and L. A. Cupples, "Bias due to missing exposure data using complete-case analysis in the proportional hazards regression model," *Statistics in Medicine*, vol. 22, no. 4, pp. 545–557, 2003.

[39] R. S. Pilla and B. G. Lindsay, "Alternative EM Methods for Non-parametric Finite Mixture Models," *Biometrika*, vol. 88, no. 2, pp. 535–550, 2001, Retrieved from http://www.jstor.org/stable/2673498.

[40] A. Allan, "Land Use Planning and its Role in Transforming the Adelaide-Gawler Line into a Transit Corridor of Connected Transit Oriented Developments," in *Paper presented at the 34th Australasian Transport Research Forum (ATRF) Proceedings*, Adelaide, Australia, 2011, http://www.worldtransitresearch.info/research/4335/.

[41] Australian Associated Motor Insurers Limited, "Adelaide's most dangerous accident hotspots revealed," 2015, Retrieved from https://www.aami.com.au/aami-answers/press-releases/adelaide-most-dangerous-accident-hotspots-revealed.html.

[42] R. O. Mujalli and J. De Oña, "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks," *Journal of Safety Research*, vol. 42, no. 5, pp. 317–326, 2011.

[43] A. Gregoriades and K. C. Mouskos, "Black spots identification through a bayesian networks quantification of accident risk index," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 28–43, 2013.

[44] R. O. Mujalli, G. López, and L. Garach, "Bayes classifiers for imbalanced traffic accidents datasets," in *Accident Analysis Prevention, (Supplement C)*, vol. 88, pp. 37–51, URL https, 2016.

[45] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," in *Accident Analysis & Prevention, 108 (Supplement C)*, pp. 27–36, 2017.

[46] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, "Learning Bayesian networks from data: an information-theory based approach," *Artificial Intelligence*, vol. 137, no. 1-2, pp. 43–90, 2002.

[47] E. Desapriya, I. Pike, and S. Babul, "Public attitudes, epidemiology and consequences of drinking and driving in british columbia," *IATSS Research*, vol. 30, no. 1, pp. 101–110, 2006.

[48] R. Hingson and M. Winter, "Epidemiology and consequences of drinking and driving," *Alcohol Research and Health*, vol. 27, no. 1, pp. 63–78, 2003, https://pubs.niaaa.nih.gov/publications/arh27-1/63-78.htm.

[49] Transport Canada's Motor Vehicle Safety Directorate, "Road Safety in Canada," 2011, Retrieved from https://www.tc.gc.ca/eng/motorvehiclesafety/tp-tp15145-1201.htm.

[50] R. Tay and S. M. Rifaat, "Factors contributing to the severity of intersection crashes," *Journal of Advanced Transportation*, vol. 41, no. 3, pp. 244–265, 2007.

[51] E. Hoareau, N. Candappa, and B. Corben, "Intersection Safety: Meeting Victoria's Intersection Challenge," *Stage 1 - Task 1: Problem Definition*, vol. 1, 2011, http://www.monash.edu/muarc/research/our-publications/muarc316a.

[52] K. Keay and I. Simmonds, "Road accidents and rainfall in a large Australian city," *Accident Analysis & Prevention*, vol. 38, no. 3, pp. 445–454, 2006.

[53] L. Symons and A. Perry, "Predicting road hazards caused by rain, freezing rain and wet surfaces and the role of weather radar," *Meteorological Applications*, vol. 4, no. 1, pp. 17–21, 1997.

[54] A. M. Rizwan, L. Y. C. Dennis, and C. Liu, "A review on the generation, determination and mitigation of Urban Heat Island," *Journal of Environmental Sciences*, vol. 20, no. 1, pp. 120–128, 2008.

[55] N. Earl, I. Simmonds, and N. Tapper, "Weekly cycles in peak time temperatures and urban heat island intensity," *Environmental Research Letters*, vol. 11, no. 7, Article ID 074003, p. 10, 2016.

[56] S. K. Jusuf, N. H. Wong, E. Hagen, R. Anggoro, and Y. Hong, "The influence of land use on the urban heat island in Singapore," *Habitat International*, vol. 31, no. 2, pp. 232–242, 2007.

[57] D. E. Parker, "Urban heat island effects on estimates of observed climate change," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 1, no. 1, pp. 123–133, 2010.

[58] M. Hosseinpour, A. S. Yahaya, and A. F. Sadullah, *Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads. Accident Analysis Prevention*, vol. 62, pp. 209–222, 2014.

[59] G. H. Bham, B. S. Javvadi, and U. R. R. Manepalli, "Multinomial logistic regression model for single-vehicle and multivehicle collisions on urban u.s. highways in arkansas," *Journal of Transportation Engineering*, vol. 138, no. 6, pp. 786–797, 2012.

[60] M. Rizzi, J. Strandroth, and C. Tingvall, "The effectiveness of antilock brake systems on motorcycles in reducing real-life crashes and injuries.," *Traffic Injury Prevention*, vol. 10, no. 5, pp. 479–487, 2009.


Journal of Engineering
The Scientific World Journal
International Journal of Rotating Machinery
Journal of Sensors
International Journal of Distributed Sensor Networks
Advances in Civil Engineering
Hindawi
Submit your manuscripts at
https://www.hindawi.com
Journal of Control Science and Engineering
Journal of Robotics
Journal of Electrical and Computer Engineering
Advances in OptoElectronics
VLSI Design
International Journal of Navigation and Observation
Modelling & Simulation in Engineering
International Journal of Aerospace Engineering
International Journal of Chemical Engineering
International Journal of Antennas and Propagation
Active and Passive Electronic Components
Shock and Vibration
Advances in Acoustics and Vibration